CRFE 1/21 03p

Access DB# ~~H2340~~

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: David Fox Examiner # : 65461 Date: 1/15/04

Art Unit: 1638 Phone Number 272-0795 Serial Number: 09/936,869

Mail Box and Bldg/Room Location: REM Results Format Preferred (circle): PAPER DISK E-MAIL

REM 2C18 2C09

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: 3/99

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

please do a SEQUENCE SEARCH

for: SEQ ID NO:1 } promoter variants of a single gene
SEQ ID NO:2 }

1	NA	1320
2	|	1401
3	|	909
5	|	1012

SEQ ID NO:3 - protein encoded by gene associated with above promoters

and SEQ ID NO:5 - coding sequence to be attached to above promoters

THANK YOU

****

STAFF USE ONLY	**Type of Search**	**Vendors and cost where applicable**
Searcher: _____	NA Sequence (#) 4	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: 1/21	Bibliographic _____	Dr.Link _____
Date Completed: 1/22	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: 12	Fulltext _____	Sequence Systems CAS
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: 12	Other _____	Other (specify) _____

PTO-1590 (8-01)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 4813.93 Seconds
(without alignments)
11217.584 Million cell updates/sec

Title: US-09-936-869-1
Perfect score: 1320
Sequence: 1 aaactcgacagcaaatatga.....tgttgctgttgcgcccattg 1320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		275.8	20.9	1726	8	AF006210	AF006210 Pinus res
2		201.8	15.3	1657	8	PMU69482	U69482 Picea maria
3		170.6	12.9	1753	8	PADAL2	X79280 P.abies dal
4		146	11.1	1642	8	PMU46582	U46582 Picea maria
5		119.4	9.0	1555	8	PMU69483	U69483 Picea maria
6	c	73.4	5.6	7218	6	I66494	I66494 Sequence 14
7	c	62.8	4.8	197271	2	AC121606	AC121606 Mus muscu
8		61.6	4.7	146695	2	AC119185	AC119185 Mus muscu
9		61	4.6	7218	6	I66494	I66494 Sequence 14
10	c	57.6	4.4	147708	2	AC129882	AC129882 Ornithorh
11		55.8	4.2	210677	2	AC074305	AC074305 Mus muscu
12	c	55	4.2	199033	2	AC134879	AC134879 Homo sapi
13		53.8	4.1	126611	10	AL928592	AL928592 Mouse DNA
14		53.2	4.0	181093	2	BX004853	BX004853 Danio rer
15	c	53	4.0	210483	2	AC111508	AC111508 Rattus no
16	c	53	4.0	222431	2	AC106170	AC106170 Rattus no
17	c	53	4.0	245412	2	AC128428	AC128428 Rattus no
18	c	52	3.9	174929	2	AC134790	AC134790 Mus muscu
19	c	51.4	3.9	216905	2	AC091332	AC091332 Mus muscu
20		51.4	3.9	235902	2	AC139195	AC139195 Mus muscu
21	c	51.2	3.9	219342	10	AL732443	AL732443 Mouse DNA
22	c	51	3.9	208244	2	AC122545	AC122545 Mus muscu
23	c	51	3.9	231443	2	AC111050	AC111050 Mus muscu
24	c	50.8	3.8	213150	2	AC107642	AC107642 Mus muscu
25	c	50.8	3.8	223940	2	AC087567	AC087567 Mus muscu
26		50.6	3.8	246611	2	AC111404	AC111404 Rattus no
27		50.2	3.8	166489	2	AC116673	AC116673 Mus muscu
28	c	50.2	3.8	199996	2	AC102649	AC102649 Mus muscu
29	c	49.8	3.8	182649	2	AC023033	AC023033 Homo sapi
30	c	49.8	3.8	250029	3	AE014820	AE014820 Plasmodiu
31	c	49.8	3.8	252988	2	AC140457	AC140457 Mus muscu
32		49.8	3.8	260967	2	AC091771	AC091771 Mus muscu
33		49.6	3.8	29221	6	AX646443	AX646443 Sequence
34		49.6	3.8	29221	9	AB065699	AB065699 Homo sapi
35	c	49.6	3.8	127425	9	AC004976	AC004976 Homo sapi
36		49.6	3.8	141819	5	AL935063	AL935063 Zebrafish
37	c	49.6	3.8	176733	2	AC115006	AC115006 Mus muscu
38	c	49.6	3.8	219367	2	AC127288	AC127288 Mus muscu
39	c	49.6	3.8	285389	2	AC096348	AC096348 Rattus no
40		49.4	3.7	110000	2	BX119988_1	Continuation (2 of
41	c	49.4	3.7	110000	2	BX119993_1	Continuation (2 of
42	c	49.4	3.7	183853	9	AP003098	AP003098 Homo sapi
43		49.4	3.7	188498	2	AC009546	AC009546 Homo sapi
44	c	49.4	3.7	266530	2	AC127680	AC127680 Mus muscu
45	c	49.2	3.7	847	11	BV076803	BV076803 S212P6027

ALIGNMENTS

RESULT 1
AF006210
LOCUS AF006210 1726 bp mRNA linear PLN 05-JAN-1999
DEFINITION Pinus resinosa MADS box transcription factor mRNA, complete cds.
ACCESSION AF006210
VERSION AF006210.1 GI:4101709
KEYWORDS .
SOURCE Pinus resinosa
ORGANISM Pinus resinosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 1726)
AUTHORS Liu,J.-J. and Podila,G.K.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1997) Department of Biological Sciences, Michigan

Technological University, Houghton, MI 49931, USA

```
FEATURES             Location/Qualifiers
     source          1. .1726
                     /organism="Pinus resinosa"
                     /mol_type="mRNA"
                     /db_xref="taxon:54921"
                     /sex="female"
                     /dev_stage="immature cones"
                     /note="Lambda TriplEx cDNA library"
     CDS             475. .1143
                     /codon_start=1
                     /product="MADS box transcription factor"
                     /protein_id="AAD01266.1"
                     /db_xref="GI:4101710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGISRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      527 a    346 c    382 g    471 t
ORIGIN
```

Query Match 20.9%; Score 275.8; DB 8; Length 1726;
Best Local Similarity 94.3%; Pred. No. 1.1e-56;
Matches 297; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

```
Qy        791 GACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGC 61

Qy        851 TGGATCATCATCATTACCATCATCGCCA-TCCCCACCATCATCATCATGATGGTATCTCT 909
              |||||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db         62 TGGATCATCATCATTACCATCATCGCGAGTCCCCACCATCATCATCATGATGGTATCTCT 121

Qy        910 ATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATC 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ATCTCTCCCTGGCAATCGATCGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATC 181

Qy        970 AACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATC 1029
              |||||||||||||||||||| |||| ||  ||||||||||||||||||||||||||||
Db        182 AACCTACCCGAAAAAACAACCTGAGTCAGCCCTGCTAATCTTGCTTATAAATCTCTTATC 241

Qy       1030 CACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTG 1089
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||| |
Db        242 CACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAACAGCGCCCGGATTGGCCGCG 301

Qy       1090 TTCTTAGATTTTCAG 1104
              ||||||||||||  ||
Db        302 TTCTTAGATTTTTAG 316
```

```
RESULT 2
PMU69482
LOCUS       PMU69482                1657 bp    mRNA    linear   PLN 24-JAN-2000
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SAG1a
            mRNA, complete cds.
ACCESSION   U69482
VERSION     U69482.1  GI:4033720
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1657)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1657)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1657
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             457. .1125
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SAG1a"
                     /protein_id="AAC97157.1"
                     /db_xref="GI:4033721"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      497 a    336 c    362 g    462 t
ORIGIN
```

Query Match 15.3%; Score 201.8; DB 8; Length 1657;
Best Local Similarity 88.7%; Pred. No. 1.4e-38;
Matches 267; Conservative 0; Mismatches 22; Indels 12; Gaps 4;

```
Qy        805 CGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCAT 864
              ||| |||||||||||||||||||||| ||||||| |||  ||||||||||||||||||||
Db          1 CGAACGACCGACCGCTCATCTTCTCCTTCTGCACAA--TTTTCTGCTGGATCATCATCAT 58

Qy        865 TACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAA 924
              || |||||||||||||||||||||||||||||||| |   ||||||||  ||||
Db         59 TATCATCATCGCCATCCCCACCATCATCATCATGATAG------TATCTCTCTTCGGCAG 112

Qy        925 TCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAAA 984
              || || ||||| |||||||||||||| ||||||||  ||||||||||||||||||||||
Db        113 TCAATCGTAGATGAAAGGAAGAGGGACGGGGCATA--TATTGATCAACCTACCCGAAAAA 170

Qy        985 ACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATT 1044
              ||||  ||||||||||  ||||| ||||||||||||||||||||||||||||||||||||
Db        171 ACAACCTGATCAGCC--TGCTAATCTGGCTTATAAATCTCTTATCCACTGTTCAATCATT 228

Qy       1045 CAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAG 1104
              |||||||||||||||||| |||||| ||||||||||||||||||||||||||||||  ||
Db        229 CAGGTTTCTTCCCACTTTGAAGCAACGGCGCCCGGATTGGCCGTGTTCTTAGATTTTTAG 288

Qy       1105 G 1105
              |
Db        289 G 289
```

```
RESULT 3
PADAL2
LOCUS       PADAL2                  1753 bp    mRNA    linear   PLN 28-FEB-1995
DEFINITION  P.abies dal2 mRNA.
ACCESSION   X79280
VERSION     X79280.1  GI:695687
KEYWORDS    DAL2 gene.
SOURCE      Picea abies (Norway spruce)
  ORGANISM  Picea abies
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1
  AUTHORS   Tandre,K., Albert,V.A., Sundas,A. and Engstrom,P.
  TITLE     Conifer homologues to genes that control floral development in
            angiosperms
  JOURNAL   Plant Mol. Biol. 27 (1), 69-78 (1995)
  MEDLINE   95170009
```

```
   PUBMED   7865797
REFERENCE   2  (bases 1 to 1753)
  AUTHORS   Tandre,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1994) K. Tandre, Uppsala University, Dept of
            Physiological Botany, Villvagen 6, 752 36 Uppsala, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .1753
                     /organism="Picea abies"
                     /mol_type="mRNA"
                     /db_xref="taxon:3329"
                     /clone="3ae"
                     /tissue_type="female cone"
                     /clone_lib="cDNA library spruce female; SF"
                     /dev_stage="before formation of ovule"
     gene            1. .1753
                     /gene="dal2"
     CDS             424. .1092
                     /gene="dal2"
                     /codon_start=1
                     /product="DAL2 protein"
                     /protein_id="CAA55867.1"
                     /db_xref="GI:695688"
                     /db_xref="SPTREMBL:Q40766"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGVISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
     misc_feature    427. .594
                     /gene="dal2"
                     /note="MADS box"
     misc_feature    697. .891
                     /gene="dal2"
                     /note="K box"
BASE COUNT      529 a    340 c    388 g    496 t
ORIGIN

  Query Match           12.9%;  Score 170.6;  DB 8;  Length 1753;
  Best Local Similarity   88.6%;  Pred. No. 6.3e-31;
  Matches  233;  Conservative    0;  Mismatches   19;  Indels   11;  Gaps    4;

Qy        843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
              ||||||||||||||||||||||| ||||||||||||||| |||||||||||||||| |
Db          5 TTTTCTGCTGGATCATCATCATTATCATCATCGCCATCCCC-CCATCATCATCATGATAG 63

Qy        903 TATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGT 962
              ||||||||   |||| || || |||||||||||||||| || ||||||||  |
Db         64 ------TATCTCTCTTCGGCAGTCAATCGTAGAGGAAAGGAAGAGAGACGGGGCATA--T 115

Qy        963 ATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATC 1022
              |||||||||||||||||||||||| ||||||||||     ||||| ||||||||||
Db        116 ATTGATCAACCTACCCGAAAAAACAACCTGATCAGCC--TGCTAATCTGGCTTATAAATC 173

Qy       1023 TCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATT 1082
              |||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||
Db        174 TCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTGAAGCAACGGCGCCCGGATT 233

Qy       1083 GGCCGTGTTCTTAGATTTTCAGG 1105
              |||||||||||||||||| |||
Db        234 GGCCGTGTTCTTAGATTTTTAGG 256


RESULT 4
PMU46582
LOCUS       PMU46582                1642 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcription factor SMADS42B
            mRNA, complete cds.
ACCESSION   U46582
VERSION     U46582.1  GI:4033709
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1642)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewardt,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1996) Robert G. Rutledge, Molecular Genetics and
            Tissue Culture, Petawawa National Forestry Institue, Chalk River,
            ON K0J 1J0, Canada
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             390. .1058
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcription factor
                     SMADS42B"
                     /protein_id="AAC97146.1"
                     /db_xref="GI:4033710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      485 a    321 c    366 g    470 t
ORIGIN

  Query Match           11.1%;  Score 146;  DB 8;  Length 1642;
  Best Local Similarity   86.4%;  Pred. No. 6.6e-25;
  Matches  185;  Conservative    0;  Mismatches   25;  Indels    4;  Gaps    2;

Qy        892 CATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAA 951
              |||||| || |  ||||||||  |||| || || |||||||||||||||||| ||
Db         13 CATCATCATCATGATAGTATCTCTCTTCGGCAGTCAATCGTAGAGGAAAGGAAGAGAGAC 72

Qy        952 GGGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTT 1011
              ||||||||  |||||||||||||||||||||||||| ||||||||||     |||||
Db         73 GGGGCATA--TATTGATCAACCTACCCGAAAAAACAACCTGATCAGCC--TGCTAATCTG 128

Qy       1012 GCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAG 1071
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| |
Db        129 GCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTGAAGCAACG 188

Qy       1072 GCGCCCGGATTGGCCGTGTTCTTAGATTTTCAGG 1105
              |||||||||||||||||||||||||||||| |||
Db        189 GCGCCCGGATTGGCCGTGTTCTTAGATTTTTAGG 222


RESULT 5
PMU69483
LOCUS       PMU69483                1555 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SMADS42C
            mRNA, complete cds.
ACCESSION   U69483
VERSION     U69483.1  GI:4033722
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
```

```
REFERENCE   1  (bases 1 to 1555)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1555)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1555
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             329. .997
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SMADS42C"
                     /protein_id="AAC97158.1"
                     /db_xref="GI:4033723"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      476 a    297 c    348 g    434 t
ORIGIN

  Query Match             9.0%;  Score 119.4;  DB 8;  Length 1555;
  Best Local Similarity   90.9%;  Pred. No. 2.2e-18;
  Matches  150;  Conservative    0;  Mismatches   11;  Indels    4;  Gaps    2;

Qy        941 GGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCC 1000
              ||||||| || |||||||| ||||||||||||||||||||||||||| ||||||||||
Db          1 GGAAGAGAGACGGGGCATA--TATTGATCAACCTACCCGAAAAAACAACCTGATCAGCCT 58

Qy       1001 TGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACT 1060
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         59 GCT--AATCTGGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACT 116

Qy       1061 TTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAGG 1105
              || |||||| |||||||||||||||||||||||||||||||| |||
Db        117 TTGAAGCAACGGCGCCCGGATTGGCCGTGTTCTTAGATTTTTAGG 161


RESULT 6
I66494/c
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match             5.6%;  Score 73.4;  DB 6;  Length 7218;
  Best Local Similarity   6.2%;  Pred. No. 3.9e-07;
  Matches   26;  Conservative  236;  Mismatches  157;  Indels    0;  Gaps    0;

Qy         93 CACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGC 152
              ||||||| ||    |    |    |    |    || || |  | ||||||| ::  ::
Db       1487 CACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRR 1428

Qy        153 TTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTAC 212
                ::     ::: :: : ::::::: ::::::: :::::: :   ::    ::  ::: :
Db       1427 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1368

Qy        213 ATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTA 272
              :  :::  :     :  ::::   ::: ::   :::   : : ::::::  :: ::   :
Db       1367 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1308

Qy        273 TAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGG 332
               :::::: : : :::: :::::: :::::::   ::: :  ::::: ::::: :: ::::
Db       1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248

Qy        333 CAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392
              ::::: : : : :::::: :    :: :: : : ::  :::: :::   : ::   ::
Db       1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188

Qy        393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452
                 :::    ::: : :::: :: :: :: :::::: :::::::::: :::: : :
Db       1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128

Qy        453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAG 511
              :::::: ::: :::::: : : :::: :::::: ::: ::: :::: : :: : :::::
Db       1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069


RESULT 7
AC121606/c
LOCUS       AC121606               197271 bp    DNA     linear   HTG 25-AUG-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-314H5, WORKING DRAFT
            SEQUENCE.
ACCESSION   AC121606
VERSION     AC121606.2  GI:22475583
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 25, 2002 this sequence version replaced gi:20986693.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0314H05
            ------------------------------------------------------
            -------------- Summary Statistics  -------------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
```

```
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197248 bases at least Q40
            Consensus quality: 197263 bases at least Q30
            Consensus quality: 197264 bases at least Q20
            Insert size: 212000; agarose-fp
            Insert size: 199833; sum-of-contigs
            Quality coverage: 14.60 in Q20 bases; agarose-fp
            Quality coverage: 12.64 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   197271: contig of 197271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197271
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-314H5"
     misc_feature    1. .197271
                     /note="assembly_name:Contig12"
BASE COUNT    56926 a  43271 c  42141 g  54929 t      4 others
ORIGIN

  Query Match             4.8%;  Score 62.8;  DB 2;  Length 197271;
  Best Local Similarity   52.7%;  Pred. No. 0.00014;
  Matches  136;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        245 TGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAAT 304
              ||| || |||| | ||   ||  |||| |||   | | | |||| ||||||  ||| |||
Db     152238 TGGCTTTGTGTAATAAAGGAAGCATTTGTAACTTAAAAAAAAAAAAAAAAAAAAAGAAAA 152179

Qy        305 TTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAAT 364
              |   || |  | ||  | |  |   | | ||||| | | | | ||||    |      | |
Db     152178 TCCTGTCTGAGGAGAAAGAGAGAGAGAGAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGA 152119

Qy        365 GCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGT 424
              | |  |   ||  |||   | |    ||       ||       |  | || | || |   |
Db     152118 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGA 152059

Qy        425 AAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGG 484
              |||||| |||| |||| | |   |    |||  || |||||  | |  ||| ||||
Db     152058 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAAAGAAAGAAAGGGAGA 151999

Qy        485 ACGAATAAATGAAATATA 502
              | |||  ||  |  | | |
Db     151998 AAGAAAGAAAGGGAGAAA 151981


RESULT 8
AC119185
LOCUS       AC119185               146695 bp    DNA     linear   HTG 07-MAR-2003
DEFINITION  Mus musculus clone RP24-174M19, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC119185
VERSION     AC119185.2  GI:28876071
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 146695)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-174M19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146695)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 146695)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 7, 2003 this sequence version replaced gi:20303909.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25169
                Center clone name: 174_M_19
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 144877 bases at least Q40
                Consensus quality: 145590 bases at least Q30
                Consensus quality: 145872 bases at least Q20
                Insert size: 145000; agarose-fp
                Insert size: 146095; sum-of-contigs
                Quality coverage: 8.6 in Q20 bases; agarose-fp
                Quality coverage: 8.5 in Q20 bases; sum-of-contigs
```

```
              --------------.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 7 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *        1     5643: contig of 5643 bp in length
              *     5644     5743: gap of 100 bp
              *     5744    75707: contig of 69964 bp in length
              *    75708    75807: gap of 100 bp
              *    75808    82448: contig of 6641 bp in length
              *    82449    82548: gap of 100 bp
              *    82549    97703: contig of 15155 bp in length
              *    97704    97803: gap of 100 bp
              *    97804   115886: contig of 18083 bp in length
              *   115887   115986: gap of 100 bp
              *   115987   144141: contig of 28155 bp in length
              *   144142   144241: gap of 100 bp
              *   144242   146695: contig of 2454 bp in length.
FEATURES             Location/Qualifiers
     source          1. .146695
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-174M19"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .5643
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5744. .75707
                     /note="assembly_fragment"
     misc_feature    75808. .82448
                     /note="assembly_fragment"
     misc_feature    82549. .97703
                     /note="assembly_fragment"
     misc_feature    97804. .115886
                     /note="assembly_fragment"
     misc_feature    115987. .144141
                     /note="assembly_fragment"
     misc_feature    144242. .146695
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    42577 a  32450 c  32243 g  38825 t    600 others
ORIGIN

  Query Match             4.7%;  Score 61.6;  DB 2;  Length 146695;
  Best Local Similarity   50.3%;  Pred. No. 0.00029;
  Matches  151;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        230 GGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAAT 289
              | || | |||| | | |  ||||   | ||| |||  ||      |  | || |  ||  |
Db      33712 GTGGATGGGATAGGTGGTTTTCGGAGGGGAAACCAGGAAAGGGGATAACATTTGAAATGT 33771

Qy        290 AAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAA 349
              ||| ||  |||| ||||||  |   |||   || |  |   |  |||||  | | | ||||
Db      33772 AAATAAAGAAAATATTTAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 33831

Qy        350 GAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATA 409
              | |      | |     | ||   || |||   | |    ||       ||    ||  |
Db      33832 AGGAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAA 33891

Qy        410 CGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAG 469
              | | | | |   | ||||||  |||| ||||  | |      |    |||  ||   |||
Db      33892 AGAAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 33951

Qy        470 CGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTT 529
              | |  |||   |||| | | |  |||   || |  ||||      ||| ||   | | |
Db      33952 AGAAAGAAAGAAAGGAAGGAAAGGTGATCCATGAGTCAAGATCAAGTTTACCCCCTTTCT 34011
```

```
RESULT 9
I66494
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match             4.6%;  Score 61;  DB 6;  Length 7218;
  Best Local Similarity   2.2%;  Pred. No. 0.00042;
  Matches    7;  Conservative  200;  Mismatches  110;  Indels    0;  Gaps    0;

Qy        603 GCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATC 662
              | ||  || || :: :  : : : ::  :   ::: ::: ::::    :::   ::: ::
Db       1055 GGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114

Qy        663 CTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTC 722
              :::::: :: :: :: ::: : ::: : :::  :   : : :::: :::: :  ::::::
Db       1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

Qy        723 CACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTT 782
              : :: ::: : : : :::  ::  ::: ::: ::::    :   : :: :::: :::::
Db       1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

Qy        783 TCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATT 842
              ::::::::  ::      :: ::  ::  ::  :: ::: :::::::::::: : :  ::
Db       1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294

Qy        843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
              :::::: ::   :: :: :: :: :: :: :: ::::: :: :: :: :: :  :  :
Db       1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354

Qy        903 TATCTCTATCTCTCCCT 919
              : ::::: :::::::::
Db       1355 YYYYYYYYYYYYYYYYY 1371


RESULT 10
AC129882/c
LOCUS       AC129882               147708 bp    DNA     linear   HTG 23-AUG-2002
DEFINITION  Ornithorhynchus anatinus clone CLM1-265D21, WORKING DRAFT SEQUENCE,
            4 ordered pieces.
ACCESSION   AC129882
VERSION     AC129882.2  GI:22450573
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Ornithorhynchus anatinus (platypus)
  ORGANISM  Ornithorhynchus anatinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
REFERENCE   1  (bases 1 to 147708)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
```

```
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147708)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 147708)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 23, 2002 this sequence version replaced gi:22122885.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: dwc
                 Center clone name: 265D21

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 146677 bases at least Q40
                 Consensus quality: 147063 bases at least Q30
                 Consensus quality: 147215 bases at least Q20
                 Insert size: 148000; agarose-fp
                 Insert size: 147408; sum-of-contigs
                 Quality coverage: 9.31x in Q20 bases; agarose-fp
                 Quality coverage: 9.35x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    87687: contig of 87687 bp in length
            *    87688    87787: gap of unknown length
            *    87788    90934: contig of 3147 bp in length
            *    90935    91034: gap of unknown length
            *    91035   124474: contig of 33440 bp in length
            *   124475   124574: gap of unknown length
            *   124575   147708: contig of 23134 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147708
                     /organism="Ornithorhynchus anatinus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9258"
                     /clone="CLM1-265D21"
                     /clone_lib="CLM1"
     misc_feature    1. .87687
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .34612
                     /note="clone overlaps with GenBank Accession Number
                     AC130185 clone CLM1-379P1 (center project name dwd)"
     misc_feature    74481. .147708
                     /note="clone overlaps with GenBank Accession Number
                     AC127463 clone CLM1-27L6 (center project name dwb)"
     misc_feature    87788. .90934
                     /note="assembly_fragment"
     misc_feature    91035. .124474
                     /note="assembly_fragment"
     misc_feature    124575. .147708
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    40144 a  33940 c  33579 g  39745 t    300 others
ORIGIN

  Query Match             4.4%;  Score 57.6;  DB 2;  Length 147708;
  Best Local Similarity   50.0%;  Pred. No. 0.0027;
  Matches  144;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy        226 TGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATA 285
              ||||||||  |||  |   |||     ||||||||      | ||   | |  ||  | |||
Db      88401 TGTCGGGGGGTGGGGGTAATGGAAATGTGTGAGAGAGAGAGAAAGAAAATAAAGAAAATA 88342

Qy        286 AAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACAC 345
              ||  || ||| ||| |||   |||    | |||  || | |  |  ||||| | | ||
Db      88341 AAGAAAGAAAGAAAGAAAAGAAAGAAAAGAAAGAAAAGAAAGAAAGAGAGAAAGAGAGAA 88282

Qy        346 AGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCG 405
              |||       |  | | | ||  ||  |||   | |    ||      ||
Db      88281 AGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAGAA 88222

Qy        406 AATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGG 465
              | | || | || |   | |||||| |||| |||| | |    |     | |  ||  |
Db      88221 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAGAAAGAG 88162

Qy        466 AGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
              |||  |   |||  ||| | ||   || || | | |  |  | ||| |
Db      88161 AGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAG 88114


RESULT 11
AC074305
LOCUS       AC074305               210677 bp    DNA     linear   HTG 26-JUL-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-10I2, WORKING DRAFT SEQUENCE,
            41 unordered pieces.
ACCESSION   AC074305
VERSION     AC074305.1  GI:9454521
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210677)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210677)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_10I2
            -------
```

```
Summary Statistics
Consensus quality: 157482 bases at least Q40
Consensus quality: 178209 bases at least Q30
Consensus quality: 186657 bases at least Q20
Estimated insert size: 201300; agarose-fp estimation
Estimated insert size: 206677; sum-of-contigs estimation
Quality coverage:  3.38 in Q20 bases; agarose-fp estimation
Quality coverage:  3.3 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1      1159: contig of 1159 bp in length
*     1160      1259: gap of unknown length
*     1260      2377: contig of 1118 bp in length
*     2378      2477: gap of unknown length
*     2478      3670: contig of 1193 bp in length
*     3671      3770: gap of unknown length
*     3771      5266: contig of 1496 bp in length
*     5267      5366: gap of unknown length
*     5367      6556: contig of 1190 bp in length
*     6557      6656: gap of unknown length
*     6657      7873: contig of 1217 bp in length
*     7874      7973: gap of unknown length
*     7974      9236: contig of 1263 bp in length
*     9237      9336: gap of unknown length
*     9337     10510: contig of 1174 bp in length
*    10511     10610: gap of unknown length
*    10611     12172: contig of 1562 bp in length
*    12173     12272: gap of unknown length
*    12273     14644: contig of 2372 bp in length
*    14645     14744: gap of unknown length
*    14745     16609: contig of 1865 bp in length
*    16610     16709: gap of unknown length
*    16710     18646: contig of 1937 bp in length
*    18647     18746: gap of unknown length
*    18747     20421: contig of 1675 bp in length
*    20422     20521: gap of unknown length
*    20522     23040: contig of 2519 bp in length
*    23041     23140: gap of unknown length
*    23141     25760: contig of 2620 bp in length
*    25761     25860: gap of unknown length
*    25861     28001: contig of 2141 bp in length
*    28002     28101: gap of unknown length
*    28102     31167: contig of 3066 bp in length
*    31168     31267: gap of unknown length
*    31268     32971: contig of 1704 bp in length
*    32972     33071: gap of unknown length
*    33072     35556: contig of 2485 bp in length
*    35557     35656: gap of unknown length
*    35657     39176: contig of 3520 bp in length
*    39177     39276: gap of unknown length
*    39277     41494: contig of 2218 bp in length
*    41495     41594: gap of unknown length
*    41595     45574: contig of 3980 bp in length
*    45575     45674: gap of unknown length
*    45675     49987: contig of 4313 bp in length
*    49988     50087: gap of unknown length
*    50088     53314: contig of 3227 bp in length
*    53315     53414: gap of unknown length
*    53415     60753: contig of 7339 bp in length
*    60754     60853: gap of unknown length
*    60854     64971: contig of 4118 bp in length
*    64972     65071: gap of unknown length
*    65072     70460: contig of 5389 bp in length
*    70461     70560: gap of unknown length
*    70561     77437: contig of 6877 bp in length
*    77438     77537: gap of unknown length
*    77538     86092: contig of 8555 bp in length
*    86093     86192: gap of unknown length
*    86193     91950: contig of 5758 bp in length
*    91951     92050: gap of unknown length
*    92051     97317: contig of 5267 bp in length
*    97318     97417: gap of unknown length
*    97418    103999: contig of 6582 bp in length
*   104000    104099: gap of unknown length
*   104100    115083: contig of 10984 bp in length
*   115084    115183: gap of unknown length
*   115184    122485: contig of 7302 bp in length
*   122486    122585: gap of unknown length
*   122586    131590: contig of 9005 bp in length
*   131591    131690: gap of unknown length
*   131691    143004: contig of 11314 bp in length
*   143005    143104: gap of unknown length
*   143105    153077: contig of 9973 bp in length
*   153078    153177: gap of unknown length
*   153178    164580: contig of 11403 bp in length
*   164581    164680: gap of unknown length
*   164681    175984: contig of 11304 bp in length
*   175985    176084: gap of unknown length
*   176085    189020: contig of 12936 bp in length
*   189021    189120: gap of unknown length
*   189121    210677: contig of 21557 bp in length.
FEATURES             Location/Qualifiers
     source          1..210677
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-10I2"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    65759 a  37166 c  38989 g  64234 t   4529 others
ORIGIN

  Query Match             4.2%;  Score 55.8;  DB 2;  Length 210677;
  Best Local Similarity   52.9%;  Pred. No. 0.0075;
  Matches  120;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        272 ATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGG 331
              | || ||| |  |   ||| ||||||  |||||    ||| |      ||    ||||  | |
Db       6987 AAAAAAAAAAAAGGAAAAAAAAAAATAAAAAGGGAAAGGGAATAAAAAAAAAAAAGAGAG 7046

Qy        332 GCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTG 391
               | ||||| | |  | |||| || |     || ||      ||  |||| |||    | ||   |
Db       7047 GAAGAAAAAAAAAAAGAAAAGAAAGAAAAAAAAAAAAAAGAAGGAAGAGAAAAAAAAAG 7106

Qy        392 ATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCAT 451
              |      ||       || |  | | |   | || | || | | ||| || |  ||||   |
Db       7107 AAAGGGAGAGAAAAAAAAAAAAAAAAGAAGGGGCAGGATAAAAAAAAAAAAGGGAGGAGAA 7166

Qy        452 TTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAA 498
                   |||| ||      | ||     | ||||     |    |   || ||| ||||
Db       7167 AGAAAAAAAAAAAAAAAGGAAAAAAAAAAAAAAAAAAAAAGAAAGGAAA 7213


RESULT 12
AC134879/c
LOCUS       AC134879               199033 bp    DNA     linear   HTG 23-FEB-2003
DEFINITION  Homo sapiens chromosome UNK clone RP11-295P22, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC134879
VERSION     AC134879.2  GI:28475765
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199033)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199033)
```

```
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 199033)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 23, 2003 this sequence version replaced gi:23396428.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0295P22
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 193954 bases at least Q40
            Consensus quality: 195124 bases at least Q30
            Consensus quality: 195874 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2110: contig of 2110 bp in length
            *     2111     2210: gap of unknown length
            *     2211     4639: contig of 2429 bp in length
            *     4640     4739: gap of unknown length
            *     4740    10008: contig of 5269 bp in length
            *    10009    10108: gap of unknown length
            *    10109    13836: contig of 3728 bp in length
            *    13837    13936: gap of unknown length
            *    13937    17705: contig of 3769 bp in length
            *    17706    17805: gap of unknown length
            *    17806    24927: contig of 7122 bp in length
            *    24928    25027: gap of unknown length
            *    25028    31944: contig of 6917 bp in length
            *    31945    32044: gap of unknown length
            *    32045    50505: contig of 18461 bp in length
            *    50506    50605: gap of unknown length
            *    50606    63617: contig of 13012 bp in length
            *    63618    63717: gap of unknown length
            *    63718    82874: contig of 19157 bp in length
            *    82875    82974: gap of unknown length
            *    82975   106052: contig of 23078 bp in length
            *   106053   106152: gap of unknown length
            *   106153   131178: contig of 25026 bp in length
            *   131179   131278: gap of unknown length
            *   131279   199033: contig of 67755 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199033
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-295P22"
     misc_feature    1. .2110
                     /note="assembly_name:Contig17"
     misc_feature    2211. .4639
                     /note="assembly_name:Contig18"
     misc_feature    4740. .10008
                     /note="assembly_name:Contig19"
     misc_feature    10109. .13836
                     /note="assembly_name:Contig20"
     misc_feature    13937. .17705
                     /note="assembly_name:Contig21"
     misc_feature    17806. .24927
                     /note="assembly_name:Contig22"
     misc_feature    25028. .31944
                     /note="assembly_name:Contig23"
     misc_feature    32045. .50505
                     /note="assembly_name:Contig24"
     misc_feature    50606. .63617
                     /note="assembly_name:Contig25"
     misc_feature    63718. .82874
                     /note="assembly_name:Contig26"
     misc_feature    82975. .106052
                     /note="assembly_name:Contig27"
     misc_feature    106153. .131178
                     /note="assembly_name:Contig28"
     misc_feature    131279. .199033
                     /note="assembly_name:Contig29"
BASE COUNT    57536 a  42075 c  35735 g  62476 t   1211 others
ORIGIN

  Query Match             4.2%;  Score 55;  DB 2;  Length 199033;
  Best Local Similarity   58.1%;  Pred. No. 0.012;
  Matches   97;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        178 AAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTG 237
              || | | | ||| | ||| || || ||| |  | | | |||| |  || |  || ||||
Db     154397 AATGGAATCAAATGGATTGGAATCAAGTGGAACAGAATGGAAAGGATTGGAAAGGACTTG 154338

Qy        238 GATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATA 297
              |||| ||||| | |  || | |  | | | |  | |||| | |||||| |||  |
Db     154337 AATGGATTGGGGTAGAAAGAAATGGAATCTAATGGGGTGGAATAGAAAAAATAGAAAGAA 154278

Qy        298 AAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACA 344
              ||| |||||||  |   |   || ||||   ||  | |||  | |
Db     154277 AAAGAATTTAATGGAGTGTGATGGAAACATATGGAATGGAATTGAAA 154231


RESULT 13
AL928592
LOCUS       AL928592              126611 bp    DNA     linear   ROD 22-MAY-2003
DEFINITION  Mouse DNA sequence from clone RP24-168J7 on chromosome 4, complete
            sequence.
ACCESSION   AL928592
VERSION     AL928592.13  GI:31071444
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 126611)
  AUTHORS   Giselle,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 23, 2003 this sequence version replaced gi:29125348.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------
```

```
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-168J7 is
            from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.
FEATURES             Location/Qualifiers
     source          1. .126611
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="4"
                     /clone="RP24-168J7"
                     /clone_lib="RPCI-24"
BASE COUNT    40625 a  24197 c  24348 g  37441 t
ORIGIN

  Query Match             4.1%;  Score 53.8;  DB 10;  Length 126611;
  Best Local Similarity   49.8%;  Pred. No. 0.023;
  Matches  136;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy        230 GGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAAT 289
              | || ||||   ||  ||    | | ||| |   |    ||||      |  || | ||||
Db      31351 GTGGGTTGGGGAGCAGGGCAAGGGGAGAGTATAGAGGGTTTTAGGGATAGCATTTGAAAT 31410

Qy        290 AAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAA 349
              |||  ||||||  || | |     | ||  |||  |   |  ||||| | | | ||||
Db      31411 GTAAATGAAAAAATATTTAATAAAAAATGTTCAAAAAGAAAGAAAGAAAGAAAGAAAGAA 31470

Qy        350 GAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATA 409
              | |      |  | ||  |||| ||      |   ||     ||    |  |
Db      31471 AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAA 31530

Qy        410 CGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAG 469
              || | || |   | ||||||  |||| ||||  |||     |     |||  ||| | ||
Db      31531 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGAAAGGAAGAAAGGAAGGAAGGA 31590

Qy        470 CGCACAAAATGGAGGACGAATAAATGAAATATA 502
              | | | || | || | |||  || |||| | |
Db      31591 AGAAAAGAAGGAAGAAAGAAAGAAAGAAAGAAA 31623


RESULT 14
BX004853
LOCUS       BX004853              181093 bp    DNA     linear   HTG 03-MAY-2003
DEFINITION  Danio rerio clone DKEY-13O11, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   BX004853
VERSION     BX004853.5  GI:30349944
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 181093)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 3, 2003 this sequence version replaced gi:30024305.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK13O11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 180623 bases at least Q40
            Consensus quality: 180751 bases at least Q30
            Consensus quality: 180906 bases at least Q20
            Insert size: 180993; sum-of-contigs
            Insert size: 172880; 4.9% error; agarose-fp
            Quality coverage: 6.77x in Q20 bases; sum-of-contigs Quality
            coverage: 7.52x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    95025: contig of 95025 bp in length
            *    95026    95125: gap of 100 bp
            *    95126   181093: contig of 85968 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181093
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-13O11"
                     /clone_lib="DanioKey"
     misc_feature    1. .95025
                     /note="assembly_fragment:00925"
     misc_feature    95126. .181093
                     /note="assembly_fragment:02465.0"
BASE COUNT    56902 a  34623 c  33889 g  55579 t    100 others
ORIGIN

  Query Match             4.0%;  Score 53.2;  DB 2;  Length 181093;
  Best Local Similarity   51.2%;  Pred. No. 0.032;
  Matches  124;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              | |||  | | |   | |||  | ||  |||  || |||||||| ||          || |
Db      81583 AAAAGGAATCTAAGGAAAAATCAAATGGAAACAAAGAAATAAAATAAAAGTTTAACTGAA 81642

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376
              | |||| || |  | |||| | | | |||||     |     |||    |  |   ||
Db      81643 AATGAAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA 81702

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436
              ||||  |||   ||      |||  ||  || | || |   | ||||| |||||
Db      81703 CAGACAGACAGACAGACAGACCGACAGACAAACAGAAAGAAAGAAAGAAAGAAAGG 81762

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496
              |||| | |    |   ||| ||  |||  | |  |||   || | |||  || ||
Db      81763 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 81822

Qy        497 AA 498
              ||
Db      81823 AA 81824


RESULT 15
AC111508/c
```

```
LOCUS       AC111508               210483 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-133F12, WORKING DRAFT SEQUENCE.
ACCESSION   AC111508
VERSION     AC111508.5  GI:30579329
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 210483)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwackelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210483)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 210483)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818983.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMLR
                Center clone name: CH230-133F12
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 202700 bases at least Q40
                Consensus quality: 203861 bases at least Q30
                Consensus quality: 204681 bases at least Q20
                Estimated insert size: 211532; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   210483: contig of 210483 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210483
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-133F12"
     misc_feature    2339. .3169
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH324644"
BASE COUNT    52766 a  43717 c  45649 g  63169 t   5182 others
ORIGIN

  Query Match             4.0%;  Score 53;  DB 2;  Length 210483;
  Best Local Similarity   52.0%;  Pred. No. 0.036;
  Matches  119;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
              |||||| | | ||| || ||| ||| |||   |   |   ||||  || | | | |
Db     204273 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGAAGAAAGAAGGAAAGAAAGAGAGAA 204214

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
              ||||| | | | |||  |  |   |  ||  |||  || | |  |    || |    ||
Db     204213 AGAAAGAAAGAAAGAGAAAGAGGGAAGAAAGAAGAAAGGAAAGAAAGAAGAAAGAAAGAA 204154

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
                ||    ||    |||| || |   | |||||  |||| | ||  ||||
Db     204153 AGAAAGAAAGAGAGAGAGAGAAAGAAAGAAAGAAGGAAAGAAGGAAAGGAGGAAGGAA 204094

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATA 502
```

Search completed: January 21, 2004, 22:14:15
Job time : 4822.93 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 5109.33 Seconds
(without alignments)
11217.584 Million cell updates/sec

Title: US-09-936-869-2
Perfect score: 1401
Sequence: 1 aaactcgacagcaaatatga.....ggaaattttgtgtacaaatc 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		280	20.0	1726	8	AF006210	AF006210 Pinus res
2		201.8	14.4	1657	8	PMU69482	U69482 Picea maria
3		170.6	12.2	1753	8	PADAL2	X79280 P.abies dal
4		146	10.4	1642	8	PMU46582	U46582 Picea maria
5		119.4	8.5	1555	8	PMU69483	U69483 Picea maria
6	c	73.4	5.2	7218	6	I66494	I66494 Sequence 14
7	c	62.8	4.5	197271	2	AC121606	AC121606 Mus muscu
8		61.6	4.4	146695	2	AC119185	AC119185 Mus muscu
9		61	4.4	7218	6	I66494	I66494 Sequence 14
10	c	57.6	4.1	147708	2	AC129882	AC129882 Ornithorh
11		55.8	4.0	210677	2	AC074305	AC074305 Mus muscu
12	c	55	3.9	199033	2	AC134879	AC134879 Homo sapi
13		53.8	3.8	126611	10	AL928592	AL928592 Mouse DNA
14		53.2	3.8	181093	2	BX004853	BX004853 Danio rer
15	c	53	3.8	210483	2	AC111508	AC111508 Rattus no
16	c	53	3.8	222431	2	AC106170	AC106170 Rattus no
17	c	53	3.8	245412	2	AC128428	AC128428 Rattus no
18	c	52	3.7	174929	2	AC134790	AC134790 Mus muscu
19	c	51.4	3.7	216905	2	AC091332	AC091332 Mus muscu
20		51.4	3.7	235902	2	AC139195	AC139195 Mus muscu
21	c	51.2	3.7	219342	10	AL732443	AL732443 Mouse DNA
22	c	51	3.6	208244	2	AC122545	AC122545 Mus muscu
23	c	51	3.6	231443	2	AC111050	AC111050 Mus muscu
24	c	50.8	3.6	213150	2	AC107642	AC107642 Mus muscu
25	c	50.8	3.6	223940	2	AC087567	AC087567 Mus muscu
26		50.6	3.6	246611	2	AC111404	AC111404 Rattus no
27		50.2	3.6	166489	2	AC116673	AC116673 Mus muscu
28	c	50.2	3.6	199996	2	AC102649	AC102649 Mus muscu
29	c	49.8	3.6	182649	2	AC023033	AC023033 Homo sapi
30	c	49.8	3.6	250029	3	AE014820	AE014820 Plasmodiu
31	c	49.8	3.6	252988	2	AC140457	AC140457 Mus muscu
32		49.8	3.6	260967	2	AC091771	AC091771 Mus muscu
33		49.6	3.5	29221	6	AX646443	AX646443 Sequence
34		49.6	3.5	29221	9	AB065699	AB065699 Homo sapi
35	c	49.6	3.5	127425	9	AC004976	AC004976 Homo sapi
36		49.6	3.5	141819	5	AL935063	AL935063 Zebrafish
37	c	49.6	3.5	176733	2	AC115006	AC115006 Mus muscu
38	c	49.6	3.5	219367	2	AC127288	AC127288 Mus muscu
39	c	49.6	3.5	285389	2	AC096348	AC096348 Rattus no
40		49.4	3.5	110000	2	BX119988_1	Continuation (2 of
41	c	49.4	3.5	110000	2	BX119993_1	Continuation (2 of
42	c	49.4	3.5	183853	9	AP003098	AP003098 Homo sapi
43		49.4	3.5	188498	2	AC009546	AC009546 Homo sapi
44	c	49.4	3.5	266530	2	AC127680	AC127680 Mus muscu
45	c	49.2	3.5	847	11	BV076803	BV076803 S212P6027

ALIGNMENTS

RESULT 1
AF006210
LOCUS AF006210 1726 bp mRNA linear PLN 05-JAN-1999
DEFINITION Pinus resinosa MADS box transcription factor mRNA, complete cds.
ACCESSION AF006210
VERSION AF006210.1 GI:4101709
KEYWORDS .
SOURCE Pinus resinosa
ORGANISM Pinus resinosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 1726)
AUTHORS Liu,J.-J. and Podila,G.K.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1997) Department of Biological Sciences, Michigan

```
            Technological University, Houghton, MI 49931, USA
FEATURES             Location/Qualifiers
     source          1. .1726
                     /organism="Pinus resinosa"
                     /mol_type="mRNA"
                     /db_xref="taxon:54921"
                     /sex="female"
                     /dev_stage="immature cones"
                     /note="Lambda TriplEx cDNA library"
     CDS             475. .1143
                     /codon_start=1
                     /product="MADS box transcription factor"
                     /protein_id="AAD01266.1"
                     /db_xref="GI:4101710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGISRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      527 a    346 c    382 g    471 t
ORIGIN

  Query Match           20.0%;  Score 280;  DB 8;  Length 1726;
  Best Local Similarity   73.9%;  Pred. No. 5.4e-58;
  Matches  452;  Conservative    0;  Mismatches   20;  Indels  140;  Gaps    2;

Qy        791 GACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGC 61

Qy        851 TGGATCATCATCATTACCATCATCGCCA-TCCCCACCATCATCATCATGATGGTATCTCT 909
              |||||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db         62 TGGATCATCATCATTACCATCATCGCGAGTCCCCACCATCATCATCATGATGGTATCTCT 121

Qy        910 ATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATC 969
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        122 ATCTCTCCCTGGCAATCGATCGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATC 181

Qy        970 AACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATC 1029
              |||||||||||||||||| ||||    ||      ||||||||||||||||||||||||
Db        182 AACCTACCCGAAAAAACAACCTGAGTCAGCCCTGCTAATCTTGCTTATAAATCTCTTATC 241

Qy       1030 CACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTG 1089
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||| |
Db        242 CACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAACAGCGCCCGGATTGGCCGCG 301

Qy       1090 TTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGTTCTGAAAAA 1149
              ||||||||||
Db        302 TTCTTAGATTT----------------------------------------------- 312

Qy       1150 GATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAAGGGGTTTGG 1209

Db        313 ------------------------------------------------------------ 312

Qy       1210 TACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCAACTGTAGGA 1269
                                          | ||||||||||||||||||||||| |||
Db        313 ----------------------------TTAGAACTCGGGCTGCGTCCAACTGTGGGA 342

Qy       1270 ACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTGCTTTTGGAC 1329
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        343 ACGAACCAGCACAAGGGCTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTGCTTTTGGAC 402

Qy       1330 TGGTATTAGTAGTTGCAGCTTTGTTTTGCATACGCTGTGAGGATCTGTGCGCGGAAATTT 1389
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TGGTAATAGTAGTTGCAGCTTTGTTTTGCATACGCTGTGAGGATCTGTGCGCGGAAATTT 462

Qy       1390 TGTGTACAAATC 1401
              ||||||||||||
Db        463 TGTGTACAAATC 474


RESULT 2
PMU69482
```

```
LOCUS       PMU69482                1657 bp    mRNA    linear   PLN 24-JAN-2000
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SAG1a
            mRNA, complete cds.
ACCESSION   U69482
VERSION     U69482.1  GI:4033720
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1657)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1657)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1657
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             457. .1125
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SAG1a"
                     /protein_id="AAC97157.1"
                     /db_xref="GI:4033721"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      497 a    336 c    362 g    462 t
ORIGIN

  Query Match           14.4%;  Score 201.8;  DB 8;  Length 1657;
  Best Local Similarity   88.7%;  Pred. No. 9.4e-39;
  Matches  267;  Conservative    0;  Mismatches   22;  Indels   12;  Gaps    4;

Qy        805 CGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCAT 864
              ||| |||||||||||||||||| ||||||| |||  ||||||||||||||||||||||
Db          1 CGAACGACCGACCGCTCATCTTCTCCTTCTGCACAA--TTTTCTGCTGGATCATCATCAT 58

Qy        865 TACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAA 924
              || |||||||||||||||||||||||||||||||||| |      |||||||| ||||
Db         59 TATCATCATCGCCATCCCCACCATCATCATCATGATAG------TATCTCTCTTCGGCAG 112

Qy        925 TCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAAA 984
              || || ||||| ||||||||||||||| |||||||| ||||||||||||||||||||||
Db        113 TCAATCGTAGATGAAAGGAAGAGGGACGGGGCATA--TATTGATCAACCTACCCGAAAAA 170

Qy        985 ACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATT 1044
              |||| |||||||||      ||||| ||||||||||||||||||||||||||||||||
Db        171 ACAACCTGATCAGCC--TGCTAATCTGGCTTATAAATCTCTTATCCACTGTTCAATCATT 228

Qy       1045 CAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAG 1104
              |||||||||||||||||| |||||| |||||||||||||||||||||||||||||||||
Db        229 CAGGTTTCTTCCCACTTTGAAGCAACGGCGCCCGGATTGGCCGTGTTCTTAGATTTTTAG 288

Qy       1105 G 1105
              |
```

```
Db        289 G 289

RESULT 3
PADAL2
LOCUS       PADAL2                  1753 bp    mRNA    linear   PLN 28-FEB-1995
DEFINITION  P.abies dal2 mRNA.
ACCESSION   X79280
VERSION     X79280.1  GI:695687
KEYWORDS    DAL2 gene.
SOURCE      Picea abies (Norway spruce)
  ORGANISM  Picea abies
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1
  AUTHORS   Tandre,K., Albert,V.A., Sundas,A. and Engstrom,P.
  TITLE     Conifer homologues to genes that control floral development in
            angiosperms
  JOURNAL   Plant Mol. Biol. 27 (1), 69-78 (1995)
  MEDLINE   95170009
   PUBMED   7865797
REFERENCE   2  (bases 1 to 1753)
  AUTHORS   Tandre,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1994) K. Tandre, Uppsala University, Dept of
            Physiological Botany, Villvagen 6, 752 36 Uppsala, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .1753
                     /organism="Picea abies"
                     /mol_type="mRNA"
                     /db_xref="taxon:3329"
                     /clone="3ae"
                     /tissue_type="female cone"
                     /clone_lib="cDNA library spruce female; SF"
                     /dev_stage="before formation of ovule"
     gene            1. .1753
                     /gene="dal2"
     CDS             424. .1092
                     /gene="dal2"
                     /codon_start=1
                     /product="DAL2 protein"
                     /protein_id="CAA55867.1"
                     /db_xref="GI:695688"
                     /db_xref="SPTREMBL:Q40766"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGVISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
     misc_feature    427. .594
                     /gene="dal2"
                     /note="MADS box"
     misc_feature    697. .891
                     /gene="dal2"
                     /note="K box"
BASE COUNT      529 a    340 c    388 g    496 t
ORIGIN

  Query Match             12.2%;  Score 170.6;  DB 8;  Length 1753;
  Best Local Similarity   88.6%;  Pred. No. 4.5e-31;
  Matches  233;  Conservative    0;  Mismatches   19;  Indels   11;  Gaps    4;

Qy        843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
              |||||||||||||||||||||| ||||||||||||||| |||||||||||||||| |
Db          5 TTTTCTGCTGGATCATCATCATTATCATCATCGCCATCCCC-CCATCATCATCATGATAG 63

Qy        903 TATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGT 962
              ||||||||   |||| || || |||||||||||||||| || ||||||||  |
Db         64 ------TATCTCTCTTCGGCAGTCAATCGTAGAGGAAAGGAAGAGAGACGGGGCATA--T 115

Qy        963 ATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATC 1022
              ||||||||||||||||||||||||| ||||||||||     ||||| ||||||||||
Db        116 ATTGATCAACCTACCCGAAAAAACAACCTGATCAGCC--TGCTAATCTGGCTTATAAATC 173

Qy       1023 TCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGATT 1082
              |||||||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||
Db        174 TCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTGAAGCAACGGCGCCCGGATT 233

Qy       1083 GGCCGTGTTCTTAGATTTTCAGG 1105
              ||||||||||||||||||| |||
Db        234 GGCCGTGTTCTTAGATTTTTAGG 256


RESULT 4
PMU46582
LOCUS       PMU46582                1642 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcription factor SMADS42B
            mRNA, complete cds.
ACCESSION   U46582
VERSION     U46582.1  GI:4033709
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1642)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewardt,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1996) Robert G. Rutledge, Molecular Genetics and
            Tissue Culture, Petawawa National Forestry Institue, Chalk River,
            ON K0J 1J0, Canada
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             390. .1058
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcription factor
                     SMADS42B"
                     /protein_id="AAC97146.1"
                     /db_xref="GI:4033710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      485 a    321 c    366 g    470 t
ORIGIN

  Query Match             10.4%;  Score 146;  DB 8;  Length 1642;
  Best Local Similarity   86.4%;  Pred. No. 5e-25;
  Matches  185;  Conservative    0;  Mismatches   25;  Indels    4;  Gaps    2;

Qy        892 CATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAA 951
              |||||| ||  |  ||||||||   |||| || || |||||||||||||||||| ||
Db         13 CATCATCATCATGATAGTATCTCTCTTCGGCAGTCAATCGTAGAGGAAAGGAAGAGAGAC 72

Qy        952 GGGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTT 1011
              ||||||||  |||||||||||||||||||||||||||| ||||||||||     |||||
Db         73 GGGGCATA--TATTGATCAACCTACCCGAAAAAACAACCTGATCAGCC--TGCTAATCTG 128

Qy       1012 GCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAG 1071
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| |
Db      129 GCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTGAAGCAACG 188

Qy     1072 GCGCCCGGATTGGCCGTGTTCTTAGATTTTCAGG 1105
            |||||||||||||||||||||||||||||| |||
Db      189 GCGCCCGGATTGGCCGTGTTCTTAGATTTTTAGG 222


RESULT 5
PMU69483
LOCUS       PMU69483                1555 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SMADS42C
            mRNA, complete cds.
ACCESSION   U69483
VERSION     U69483.1  GI:4033722
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1555)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1555)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1..1555
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             329..997
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SMADS42C"
                     /protein_id="AAC97158.1"
                     /db_xref="GI:4033723"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      476 a    297 c    348 g    434 t
ORIGIN

  Query Match             8.5%;  Score 119.4;  DB 8;  Length 1555;
  Best Local Similarity   90.9%;  Pred. No. 1.8e-18;
  Matches  150;  Conservative    0;  Mismatches   11;  Indels    4;  Gaps    2;

Qy      941 GGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCC 1000
            ||||||| || ||||||||  |||||||||||||||||||||||||| ||||||||||
Db        1 GGAAGAGAGACGGGGCATA--TATTGATCAACCTACCCGAAAAAACAACCTGATCAGCCT 58

Qy     1001 TGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACT 1060
                  ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       59 GCT--AATCTGGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACT 116

Qy     1061 TTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAGG 1105
            || |||||  ||||||||||||||||||||||||||||||| |||
Db      117 TTGAAGCAACGGCGCCCGGATTGGCCGTGTTCTTAGATTTTTAGG 161
```

```
RESULT 6
I66494/c
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1..7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match             5.2%;  Score 73.4;  DB 6;  Length 7218;
  Best Local Similarity   6.2%;  Pred. No. 3.6e-07;
  Matches   26;  Conservative  236;  Mismatches  157;  Indels    0;  Gaps    0;

Qy       93 CACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGC 152
            |||||||  ||      |       |      |    || |  |   |||||||| ::  ::
Db     1487 CACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRR 1428

Qy      153 TTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTAC 212
              ::     ::: :: : :::::: ::::::: :::::: :   ::     ::  ::: :
Db     1427 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1368

Qy      213 ATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTA 272
             :  :::  :      :  ::::   ::: ::    :::    : : :::::: :: ::   :
Db     1367 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1308

Qy      273 TAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGG 332
              :::::: : : :::: ::::: ::::::       ::: :   ::::: ::::: :: ::::
Db     1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248

Qy      333 CAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392
             ::::: : : : ::::::: :     :: :: : : ::  :::: :::    : ::   ::
Db     1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188

Qy      393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452
                 :::    ::: : :::: :: :: :: :::::: ::::::::: :::: : :
Db     1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128

Qy      453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAG 511
              ::::: ::: :::::: : : :::: :::::: ::: ::: :::: : :: : :::::
Db     1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069


RESULT 7
AC121606/c
LOCUS       AC121606              197271 bp    DNA     linear   HTG 25-AUG-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-314H5, WORKING DRAFT
            SEQUENCE.
ACCESSION   AC121606
VERSION     AC121606.2  GI:22475583
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (20-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 197271)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Aug 25, 2002 this sequence version replaced gi:20986693.

            ---------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0314H05
            -----------------------------------------------------
            -------------- Summary Statistics -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197248 bases at least Q40
            Consensus quality: 197263 bases at least Q30
            Consensus quality: 197264 bases at least Q20
            Insert size: 212000; agarose-fp
            Insert size: 199833; sum-of-contigs
            Quality coverage: 14.60 in Q20 bases; agarose-fp
            Quality coverage: 12.64 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1  197271: contig of 197271 bp in length.
FEATURES             Location/Qualifiers
     source          1..197271
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-314H5"
     misc_feature    1..197271
                     /note="assembly_name:Contig12"
BASE COUNT    56926 a  43271 c  42141 g  54929 t      4 others
ORIGIN

  Query Match             4.5%;  Score 62.8;  DB 2;  Length 197271;
  Best Local Similarity   52.7%;  Pred. No. 0.00014;
  Matches  136;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        245 TGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAAT 304
              |||  ||  ||||  |  ||      ||   ||||  |||     |  |  |  ||||  ||||||  |||  |||
Db     152238 TGGCTTTGTGTAATAAAGGAAGCATTTGTAACTTAAAAAAAAAAAAAAAAAAAAAAGAAAA 152179

Qy        305 TTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAAT 364
              |     ||  |  |  ||  |  |   |     |  |  ||||||  |  |  |  ||||     |      |  |
Db     152178 TCCTGTCTGAGGAGAAAGAGAGAGAGAGAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGA 152119

Qy        365 GCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGT 424
              |  |    |   ||   |||          |      ||          |     |  |  |  |  |  |    |
Db     152118 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGA 152059

Qy        425 AAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGG 484
              ||||||  ||||  ||||  |  |     |          |||  ||  |||||  |  |   |||  ||||
Db     152058 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGGAGAAAGAAAGAAAGGGAGA 151999

Qy        485 ACGAATAAATGAAATATA 502
              |  |||  ||  |  |  |
Db     151998 AAGAAAGAAAGGGAGAAA 151981


RESULT 8
AC119185
LOCUS       AC119185              146695 bp    DNA     linear   HTG 07-MAR-2003
DEFINITION  Mus musculus clone RP24-174M19, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC119185
VERSION     AC119185.2  GI:28876071
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 146695)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-174M19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146695)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 146695)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 7, 2003 this sequence version replaced gi:20303909.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          ------------- Project Information
              Center project name: L25169
              Center clone name: 174_M_19
          ------------- Summary Statistics
              Sequencing vector: Plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 144877 bases at least Q40
              Consensus quality: 145590 bases at least Q30
              Consensus quality: 145872 bases at least Q20
              Insert size: 145000; agarose-fp
              Insert size: 146095; sum-of-contigs
              Quality coverage: 8.6 in Q20 bases; agarose-fp
              Quality coverage: 8.5 in Q20 bases; sum-of-contigs
          -------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     5643: contig of 5643 bp in length
          *     5644     5743: gap of 100 bp
          *     5744    75707: contig of 69964 bp in length
          *    75708    75807: gap of 100 bp
          *    75808    82448: contig of 6641 bp in length
          *    82449    82548: gap of 100 bp
          *    82549    97703: contig of 15155 bp in length
          *    97704    97803: gap of 100 bp
          *    97804   115886: contig of 18083 bp in length
          *   115887   115986: gap of 100 bp
          *   115987   144141: contig of 28155 bp in length
          *   144142   144241: gap of 100 bp
          *   144242   146695: contig of 2454 bp in length.
FEATURES             Location/Qualifiers
     source          1. .146695
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-174M19"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .5643
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5744. .75707
                     /note="assembly_fragment"
     misc_feature    75808. .82448
                     /note="assembly_fragment"
     misc_feature    82549. .97703
                     /note="assembly_fragment"
     misc_feature    97804. .115886
                     /note="assembly_fragment"
     misc_feature    115987. .144141
                     /note="assembly_fragment"
     misc_feature    144242. .146695
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    42577 a  32450 c  32243 g  38825 t    600 others
ORIGIN

  Query Match             4.4%;  Score 61.6;  DB 2;  Length 146695;
  Best Local Similarity   50.3%;  Pred. No. 0.00028;
  Matches  151;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        230 GGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAAT 289
              | || | |||| | | | ||||   | ||| |||  ||    |  | ||  |  ||  |
Db      33712 GTGGATGGGATAGGTGGTTTTCGGAGGGGAAACCAGGAAAGGGGATAACATTTGAAATGT 33771

Qy        290 AAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAA 349
              ||| ||  |||| |||||  |  |||  ||  |  |  |  |||||  | | | ||||
Db      33772 AAATAAAGAAAATATTTAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 33831

Qy        350 GAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATA 409
              | |     | |   | ||  ||  |||  | |   ||     ||   ||  | |
Db      33832 AGGAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAA 33891

Qy        410 CGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAG 469
              || |  | |  | ||||||  |||| ||||  | |     |  |||  ||  |||
Db      33892 AGAAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA 33951

Qy        470 CGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTT 529
              | |  |||   |||| | | | |||   ||  |  ||||     ||| ||   | | |
Db      33952 AGAAAGAAAGAAAGGAAGGAAAGGTGATCCATGAGTCAAGATCAAGTTTACCCCCTTTCT 34011


RESULT 9
I66494
LOCUS       I66494                 7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match             4.4%;  Score 61;  DB 6;  Length 7218;
  Best Local Similarity   2.2%;  Pred. No. 0.00041;
  Matches    7;  Conservative  200;  Mismatches  110;  Indels    0;  Gaps    0;

Qy        603 GCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATC 662
              | ||  || || :: :  : : : ::  :   ::: ::: ::::    :::   ::: ::
Db       1055 GGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114

Qy        663 CTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTC 722
              :::::: :: :: :: ::: : ::: : :::  :   : : :::: :::: :  ::::::
Db       1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

Qy        723 CACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTT 782
              : :: ::: : : : :::  ::  ::: ::: :::::  :  : :: :::: :::::
Db       1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

Qy        783 TCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATT 842
              :::::::  ::          :: ::  ::  ::  :: ::: :::::::::::: : :  ::
Db       1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294

Qy        843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
              :::::: ::   :: :: :: :: :: :: :: :: :: ::::: :: :: :: :: : :
Db       1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354

Qy        903 TATCTCTATCTCTCCCT 919
              : ::::: :::::::::
Db       1355 YYYYYYYYYYYYYYYYY 1371


RESULT 10
AC129882/c
```

```
LOCUS       AC129882              147708 bp    DNA     linear   HTG 23-AUG-2002
DEFINITION  Ornithorhynchus anatinus clone CLM1-265D21, WORKING DRAFT SEQUENCE,
            4 ordered pieces.
ACCESSION   AC129882
VERSION     AC129882.2  GI:22450573
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Ornithorhynchus anatinus (platypus)
  ORGANISM  Ornithorhynchus anatinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
REFERENCE   1  (bases 1 to 147708)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147708)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 147708)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 23, 2002 this sequence version replaced gi:22122885.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: dwc
                 Center clone name: 265D21

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 146677 bases at least Q40
                 Consensus quality: 147063 bases at least Q30
                 Consensus quality: 147215 bases at least Q20
                 Insert size: 148000; agarose-fp
                 Insert size: 147408; sum-of-contigs
                 Quality coverage: 9.31x in Q20 bases; agarose-fp
                 Quality coverage: 9.35x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    87687: contig of 87687 bp in length
            *   87688    87787: gap of unknown length
            *   87788    90934: contig of 3147 bp in length
            *   90935    91034: gap of unknown length
            *   91035   124474: contig of 33440 bp in length
            *  124475   124574: gap of unknown length
            *  124575   147708: contig of 23134 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147708
                     /organism="Ornithorhynchus anatinus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9258"
                     /clone="CLM1-265D21"
                     /clone_lib="CLM1"
     misc_feature    1. .87687
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .34612
                     /note="clone overlaps with GenBank Accession Number
                     AC130185 clone CLM1-379P1 (center project name dwd)"
     misc_feature    74481. .147708
                     /note="clone overlaps with GenBank Accession Number
                     AC127463 clone CLM1-27L6 (center project name dwb)"
     misc_feature    87788. .90934
                     /note="assembly_fragment"
     misc_feature    91035. .124474
                     /note="assembly_fragment"
     misc_feature    124575. .147708
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    40144 a  33940 c  33579 g  39745 t    300 others
ORIGIN

  Query Match             4.1%;  Score 57.6;  DB 2;  Length 147708;
  Best Local Similarity   50.0%;  Pred. No. 0.0027;
  Matches  144;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy        226 TGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATA 285
              ||||||| ||| |   |||    ||||||||     | ||   | |  ||  | |||
Db      88401 TGTCGGGGGGTGGGGGTAATGGAAATGTGTGAGAGAGAGAGAAAGAAAATAAAGAAAATA 88342

Qy        286 AAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACAC 345
              ||  || ||| ||| |||   |||    | |||  || |  |   |  ||||| | | |
Db      88341 AAGAAAGAAAGAAAGAAAAGAAAGAAAAGAAAGAAAAGAAAGAAAGAGAGAAAGAGAGAA 88282

Qy        346 AGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCG 405
              |||    |    | |  | | ||   ||  |||   | |    ||      ||
Db      88281 AGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAGAA 88222

Qy        406 AATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGG 465
              |  | || | || |   | ||||||| |||| ||||  | |    |      | |  || |
Db      88221 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAGAAAGAG 88162

Qy        466 AGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
              |||  |   |||  ||| | ||   || || | | |  |  | || |
Db      88161 AGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAG 88114


RESULT 11
AC074305
LOCUS       AC074305              210677 bp    DNA     linear   HTG 26-JUL-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-10I2, WORKING DRAFT SEQUENCE,
            41 unordered pieces.
ACCESSION   AC074305
VERSION     AC074305.1  GI:9454521
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210677)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210677)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_10I2
            -------
            Summary Statistics
            Consensus quality: 157482 bases at least Q40
            Consensus quality: 178209 bases at least Q30
            Consensus quality: 186657 bases at least Q20
            Estimated insert size: 201300; agarose-fp estimation
            Estimated insert size: 206677; sum-of-contigs estimation
            Quality coverage:  3.38 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.3 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 41 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1159: contig of 1159 bp in length
            *     1160     1259: gap of unknown length
            *     1260     2377: contig of 1118 bp in length
            *     2378     2477: gap of unknown length
            *     2478     3670: contig of 1193 bp in length
            *     3671     3770: gap of unknown length
            *     3771     5266: contig of 1496 bp in length
            *     5267     5366: gap of unknown length
            *     5367     6556: contig of 1190 bp in length
            *     6557     6656: gap of unknown length
            *     6657     7873: contig of 1217 bp in length
            *     7874     7973: gap of unknown length
            *     7974     9236: contig of 1263 bp in length
            *     9237     9336: gap of unknown length
            *     9337    10510: contig of 1174 bp in length
            *    10511    10610: gap of unknown length
            *    10611    12172: contig of 1562 bp in length
            *    12173    12272: gap of unknown length
            *    12273    14644: contig of 2372 bp in length
            *    14645    14744: gap of unknown length
            *    14745    16609: contig of 1865 bp in length
            *    16610    16709: gap of unknown length
            *    16710    18646: contig of 1937 bp in length
            *    18647    18746: gap of unknown length
            *    18747    20421: contig of 1675 bp in length
            *    20422    20521: gap of unknown length
            *    20522    23040: contig of 2519 bp in length
            *    23041    23140: gap of unknown length
            *    23141    25760: contig of 2620 bp in length
            *    25761    25860: gap of unknown length
            *    25861    28001: contig of 2141 bp in length
            *    28002    28101: gap of unknown length
            *    28102    31167: contig of 3066 bp in length
            *    31168    31267: gap of unknown length
            *    31268    32971: contig of 1704 bp in length
            *    32972    33071: gap of unknown length
            *    33072    35556: contig of 2485 bp in length
            *    35557    35656: gap of unknown length
            *    35657    39176: contig of 3520 bp in length
            *    39177    39276: gap of unknown length
            *    39277    41494: contig of 2218 bp in length
            *    41495    41594: gap of unknown length
            *    41595    45574: contig of 3980 bp in length
            *    45575    45674: gap of unknown length
            *    45675    49987: contig of 4313 bp in length
            *    49988    50087: gap of unknown length
            *    50088    53314: contig of 3227 bp in length
            *    53315    53414: gap of unknown length
            *    53415    60753: contig of 7339 bp in length
            *    60754    60853: gap of unknown length
            *    60854    64971: contig of 4118 bp in length
            *    64972    65071: gap of unknown length
            *    65072    70460: contig of 5389 bp in length
            *    70461    70560: gap of unknown length
            *    70561    77437: contig of 6877 bp in length
            *    77438    77537: gap of unknown length
            *    77538    86092: contig of 8555 bp in length
            *    86093    86192: gap of unknown length
            *    86193    91950: contig of 5758 bp in length
            *    91951    92050: gap of unknown length
            *    92051    97317: contig of 5267 bp in length
            *    97318    97417: gap of unknown length
            *    97418   103999: contig of 6582 bp in length
            *   104000   104099: gap of unknown length
            *   104100   115083: contig of 10984 bp in length
            *   115084   115183: gap of unknown length
            *   115184   122485: contig of 7302 bp in length
            *   122486   122585: gap of unknown length
            *   122586   131590: contig of 9005 bp in length
            *   131591   131690: gap of unknown length
            *   131691   143004: contig of 11314 bp in length
            *   143005   143104: gap of unknown length
            *   143105   153077: contig of 9973 bp in length
            *   153078   153177: gap of unknown length
            *   153178   164580: contig of 11403 bp in length
            *   164581   164680: gap of unknown length
            *   164681   175984: contig of 11304 bp in length
            *   175985   176084: gap of unknown length
            *   176085   189020: contig of 12936 bp in length
            *   189021   189120: gap of unknown length
            *   189121   210677: contig of 21557 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210677
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-10I2"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    65759 a  37166 c  38989 g  64234 t   4529 others
ORIGIN

  Query Match             4.0%;  Score 55.8;  DB 2;  Length 210677;
  Best Local Similarity   52.9%;  Pred. No. 0.0075;
  Matches  120;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        272 ATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGG 331
              | || ||| | |  ||| ||||||  |||||    ||| |    ||  |||| | | |
Db       6987 AAAAAAAAAAAAGGAAAAAAAAAAATAAAAAGGGAAAGGGAATAAAAAAAAAAAAGAGAG 7046

Qy        332 GCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTG 391
              | ||||| | | | |||| || |   || ||     ||  |||| |||   | ||   |
Db       7047 GAAGAAAAAAAAAAAGAAAAGAAAGAAAAAAAAAAAAAAAGAAGGAAGAGAAAAAAAAAG 7106

Qy        392 ATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCAT 451
              |      ||  |  | | |  | || | || | |||  || |  ||||   |
Db       7107 AAAGGGAGAGAAAAAAAAAAAAAAAGAAGGGGCAGGATAAAAAAAAAAAAGGGAGGAGAA 7166

Qy        452 TTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAA 498
                  ||||  ||   | ||   | ||||   |  |  || ||| ||||
```

```
Db       7167 AGAAAAAAAAAAAAAAAGGAAAAAAAAAAAAAAAAAAAAAGAAAGGAAA 7213


RESULT 12
AC134879/c
LOCUS       AC134879              199033 bp    DNA     linear   HTG 23-FEB-2003
DEFINITION  Homo sapiens chromosome UNK clone RP11-295P22, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC134879
VERSION     AC134879.2  GI:28475765
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199033)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199033)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 199033)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 23, 2003 this sequence version replaced gi:23396428.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0295P22
            -----------------------------------------------------
            ------------- Summary Statistics --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 193954 bases at least Q40
            Consensus quality: 195124 bases at least Q30
            Consensus quality: 195874 bases at least Q20
            -----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2110: contig of 2110 bp in length
            *     2111     2210: gap of unknown length
            *     2211     4639: contig of 2429 bp in length
            *     4640     4739: gap of unknown length
            *     4740    10008: contig of 5269 bp in length
            *    10009    10108: gap of unknown length
            *    10109    13836: contig of 3728 bp in length
            *    13837    13936: gap of unknown length
            *    13937    17705: contig of 3769 bp in length
            *    17706    17805: gap of unknown length
            *    17806    24927: contig of 7122 bp in length
            *    24928    25027: gap of unknown length
            *    25028    31944: contig of 6917 bp in length
            *    31945    32044: gap of unknown length
            *    32045    50505: contig of 18461 bp in length
            *    50506    50605: gap of unknown length
            *    50606    63617: contig of 13012 bp in length
            *    63618    63717: gap of unknown length
            *    63718    82874: contig of 19157 bp in length
            *    82875    82974: gap of unknown length
            *    82975   106052: contig of 23078 bp in length
            *   106053   106152: gap of unknown length
            *   106153   131178: contig of 25026 bp in length
            *   131179   131278: gap of unknown length
            *   131279   199033: contig of 67755 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199033
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-295P22"
     misc_feature    1. .2110
                     /note="assembly_name:Contig17"
     misc_feature    2211. .4639
                     /note="assembly_name:Contig18"
     misc_feature    4740. .10008
                     /note="assembly_name:Contig19"
     misc_feature    10109. .13836
                     /note="assembly_name:Contig20"
     misc_feature    13937. .17705
                     /note="assembly_name:Contig21"
     misc_feature    17806. .24927
                     /note="assembly_name:Contig22"
     misc_feature    25028. .31944
                     /note="assembly_name:Contig23"
     misc_feature    32045. .50505
                     /note="assembly_name:Contig24"
     misc_feature    50606. .63617
                     /note="assembly_name:Contig25"
     misc_feature    63718. .82874
                     /note="assembly_name:Contig26"
     misc_feature    82975. .106052
                     /note="assembly_name:Contig27"
     misc_feature    106153. .131178
                     /note="assembly_name:Contig28"
     misc_feature    131279. .199033
                     /note="assembly_name:Contig29"
BASE COUNT    57536 a   42075 c   35735 g   62476 t    1211 others
ORIGIN

  Query Match             3.9%;  Score 55;  DB 2;  Length 199033;
  Best Local Similarity   58.1%;  Pred. No. 0.012;
  Matches   97;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        178 AAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTG 237
              || | | | ||| | ||| || || ||| |  | | | |||| |  || |   || ||||
Db     154397 AATGGAATCAAATGGATTGGAATCAAGTGGAACAGAATGGAAAGGATTGGAAAGGACTTG 154338

Qy        238 GATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATA 297
              |||| ||||| | |   || | |  | | | |     | |||| | |||||| |||  |
Db     154337 AATGGATTGGGGTAGAAAGAAATGGAATCTAATGGGGTGGAATAGAAAAAATAGAAAGAA 154278

Qy        298 AAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACA 344
              ||| |||||||  |   |   || ||||    ||   | |||  | |
Db     154277 AAAGAATTTAATGGAGTGTGATGGAAACATATGGAATGGAATTGAAA 154231


RESULT 13
AL928592
LOCUS       AL928592              126611 bp    DNA     linear   ROD 22-MAY-2003
DEFINITION  Mouse DNA sequence from clone RP24-168J7 on chromosome 4, complete
            sequence.
ACCESSION   AL928592
VERSION     AL928592.13  GI:31071444
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 126611)
  AUTHORS   Giselle,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 23, 2003 this sequence version replaced gi:29125348.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-168J7 is
            from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.
FEATURES             Location/Qualifiers
     source          1. .126611
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="4"
                     /clone="RP24-168J7"
                     /clone_lib="RPCI-24"
BASE COUNT    40625 a  24197 c  24348 g  37441 t
ORIGIN

  Query Match             3.8%;  Score 53.8;  DB 10;  Length 126611;
  Best Local Similarity   49.8%;  Pred. No. 0.023;
  Matches  136;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy        230 GGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAAT 289
              | || ||||   ||  ||    | | ||| |   |    ||||      |  || | ||||
Db      31351 GTGGGTTGGGGAGCAGGGCAAGGGGAGAGTATAGAGGGTTTTAGGGATAGCATTTGAAAT 31410

Qy        290 AAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAA 349
              |||  ||||||  || | |      | ||  |||  |   |  |||||  | | | ||||
Db      31411 GTAAATGAAAAAATATTTAATAAAAAATGTTCAAAAAGAAAGAAAGAAAGAAAGAAAGAA 31470

Qy        350 GAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATA 409
              | |       |   | ||  |||| ||      |   ||      ||     |  |
Db      31471 AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAA 31530

Qy        410 CGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAG 469
              || | ||  |  ||||||  |||| ||||  |||     |     ||| ||| | ||
Db      31531 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGAAAGGAAGAAAGGAAGGAAGGA 31590

Qy        470 CGCACAAAATGGAGGACGAATAAATGAAATATA 502
              | | | || | || | |||  || |||| | |
```

```
Db      31591 AGAAAAGAAGGAAGAAAGAAAGAAAGAAAGAAA 31623


RESULT 14
BX004853
LOCUS       BX004853              181093 bp    DNA     linear   HTG 03-MAY-2003
DEFINITION  Danio rerio clone DKEY-13O11, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   BX004853
VERSION     BX004853.5  GI:30349944
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 181093)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 3, 2003 this sequence version replaced gi:30024305.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK13O11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 180623 bases at least Q40
            Consensus quality: 180751 bases at least Q30
            Consensus quality: 180906 bases at least Q20
            Insert size: 180993; sum-of-contigs
            Insert size: 172880; 4.9% error; agarose-fp
            Quality coverage: 6.77x in Q20 bases; sum-of-contigs Quality
            coverage: 7.52x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    95025: contig of 95025 bp in length
            *   95026    95125: gap of 100 bp
            *   95126   181093: contig of 85968 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181093
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-13O11"
                     /clone_lib="DanioKey"
     misc_feature    1. .95025
                     /note="assembly_fragment:00925"
     misc_feature    95126. .181093
                     /note="assembly_fragment:02465.0"
BASE COUNT    56902 a  34623 c  33889 g  55579 t    100 others
ORIGIN

  Query Match             3.8%;  Score 53.2;  DB 2;  Length 181093;
  Best Local Similarity   51.2%;  Pred. No. 0.033;
  Matches  124;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              | |||  | | |   | ||   | ||  |||  || |||||||| ||         || |
Db      81583 AAAAGGAATCTAAGGAAAAATCAAATGGAAACAAAGAAATAAAATAAAAGTTTAACTGAA 81642
```

```
Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376
              | |||| || |   | |||| | | | |||||      |     |||    |  |   ||
Db      81643 AATGAAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA 81702

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436
              ||||   |||     ||        |||     ||  | || |   | ||||| |||||
Db      81703 CAGACAGACAGACAGACAGACCGACAGACAAACAGAAAGAAAGAAAGAAAGAAAGAAAGG 81762

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496
              ||||  | |     |   |||  ||   |||  | |  |||   || | |||  || ||
Db      81763 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 81822

Qy        497 AA 498
              ||
Db      81823 AA 81824
```

RESULT 15
AC111508/c

```
LOCUS       AC111508               210483 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-133F12, WORKING DRAFT SEQUENCE.
ACCESSION   AC111508
VERSION     AC111508.5  GI:30579329
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 210483)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210483)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 210483)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818983.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMLR
                Center clone name: CH230-133F12
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 202700 bases at least Q40
                Consensus quality: 203861 bases at least Q30
                Consensus quality: 204681 bases at least Q20
                Estimated insert size: 211532; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   210483: contig of 210483 bp in length.
FEATURES                Location/Qualifiers
     source             1. .210483
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-133F12"
     misc_feature       2339. .3169
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:BH324644"
```

BASE COUNT 52766 a 43717 c 45649 g 63169 t 5182 others
ORIGIN

Query Match 3.8%; Score 53; DB 2; Length 210483;
Best Local Similarity 52.0%; Pred. No. 0.037;
Matches 119; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```
Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
              |||||| | | |||  || ||| ||| |||   |   |   |||| || |  | | |
Db     204273 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGAAGAAAGAAGGAAAGAAAGAGAA 204214

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
              ||||| | | | |||  |  |   |  ||  | ||  || | |      | |    ||
Db     204213 AGAAAGAAAGAAAGAGAAAGAGGGAAGAAAGAAGAAAGGAAAGAAAGAAGAAAGAAAGAA 204154

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
                ||   ||    |||| || |   | |||||  |||| | ||  ||||
Db     204153 AGAAAGAAAGAGAGAGAGAGAAAGAAAGAAAGAAGGAAAGAAGGAAAGGAGGAAGGAA 204094

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATA 502
              | |||  ||  ||||| | |  |||   || | |||  || ||||  | |
Db     204093 GGAAAGAAAGAGAGAGAGAAAGAAAGAAAGAAGGAAAGAAGGAAAGAAA 204045
```

Search completed: January 21, 2004, 22:14:19
Job time : 5113.33 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 16:52:57 ; Search time 61.4877 Seconds
(without alignments)
6525.161 Million cell updates/sec

Title: US-09-936-869-3
Perfect score: 909
Sequence: 1 atgggtcgtgggaagattga.....cgcatgttatctttctgttg 909

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	260	28.6	1219	4	US-09-410-464-10	Sequence 10, Appl
2	259.8	28.6	723	4	US-09-410-464-11	Sequence 11, Appl
3	255.6	28.1	959	3	US-09-067-800-7	Sequence 7, Appli
4	255.6	28.1	959	3	US-09-349-677-7	Sequence 7, Appli
5	254.6	28.0	1125	1	US-08-460-512-6	Sequence 6, Appli
6	248.2	27.3	714	4	US-09-410-464-15	Sequence 15, Appl
7	248.2	27.3	1159	4	US-09-410-464-14	Sequence 14, Appl
8	248.2	27.3	1297	4	US-09-433-579-1	Sequence 1, Appli
9	245	27.0	896	3	US-09-067-800-5	Sequence 5, Appli
10	245	27.0	896	3	US-09-349-677-5	Sequence 5, Appli
11	232.4	25.6	1457	1	US-08-460-512-1	Sequence 1, Appli
12	230.8	25.4	1097	1	US-08-460-512-3	Sequence 3, Appli
13	147.2	16.2	1043	2	US-08-867-087B-12	Sequence 12, Appl
14	141.6	15.6	1062	3	US-09-067-800-1	Sequence 1, Appli
15	141.6	15.6	1062	3	US-09-105-652-1	Sequence 1, Appli
16	141.6	15.6	1062	3	US-09-349-677-1	Sequence 1, Appli
17	141.6	15.6	1062	4	US-09-708-584-1	Sequence 1, Appli
18	138	15.2	1345	1	US-08-592-214A-7	Sequence 7, Appli
19	138	15.2	1345	3	US-08-659-188-7	Sequence 7, Appli
20	138	15.2	1345	3	US-08-655-227-7	Sequence 7, Appli
21	138	15.2	1345	3	US-08-655-241-7	Sequence 7, Appli
22	138	15.2	1345	3	US-09-149-976-7	Sequence 7, Appli
23	138	15.2	1345	4	US-09-398-326-7	Sequence 7, Appli
24	131.2	14.4	795	3	US-08-904-284-2	Sequence 2, Appli
25	129.2	14.2	1180	2	US-08-867-087B-16	Sequence 16, Appl
26	126.8	13.9	794	1	US-08-592-214A-3	Sequence 3, Appli
27	126.8	13.9	794	3	US-08-659-188-3	Sequence 3, Appli
28	126.8	13.9	794	3	US-08-655-227-3	Sequence 3, Appli
29	126.8	13.9	794	3	US-08-655-241-3	Sequence 3, Appli
30	126.8	13.9	794	3	US-09-149-976-3	Sequence 3, Appli
31	126.8	13.9	794	4	US-09-398-326-3	Sequence 3, Appli
32	125.2	13.8	768	1	US-08-592-214A-5	Sequence 5, Appli
33	125.2	13.8	768	3	US-08-659-188-5	Sequence 5, Appli
34	125.2	13.8	768	3	US-08-655-227-5	Sequence 5, Appli
35	125.2	13.8	768	3	US-08-655-241-5	Sequence 5, Appli
36	125.2	13.8	768	3	US-09-149-976-5	Sequence 5, Appli
37	125.2	13.8	768	4	US-09-398-326-5	Sequence 5, Appli
38	124.4	13.7	756	1	US-08-592-214A-11	Sequence 11, Appl
39	124.4	13.7	756	3	US-08-659-188-11	Sequence 11, Appl
40	124.4	13.7	756	3	US-08-655-227-11	Sequence 11, Appl
41	124.4	13.7	756	3	US-08-655-241-11	Sequence 11, Appl
42	124.4	13.7	756	3	US-09-149-976-11	Sequence 11, Appl
43	124.4	13.7	756	4	US-09-398-326-11	Sequence 11, Appl
44	124	13.6	756	1	US-08-592-214A-13	Sequence 13, Appl
45	124	13.6	756	3	US-08-659-188-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1 .
US-09-410-464-10
; Sequence 10, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 53375
; CURRENT APPLICATION NUMBER: US/09/410,464
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 09/287,700
; EARLIER FILING DATE: 1999-04-06
; EARLIER APPLICATION NUMBER: 60/080,851
; EARLIER FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1219
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(921)
US-09-410-464-10

Query Match 28.6%; Score 260; DB 4; Length 1219;
Best Local Similarity 63.7%; Pred. No. 5.5e-67;
Matches 433; Conservative 0; Mismatches 235; Indels 12; Gaps 2;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db        242 TGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 301

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              ||||||| | || | |||||| | ||||| || || |||||||| |||||||| |||| 
Db        302 TCTGCAAAAGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATG 361

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || || ||||||||||||||| ||| | ||  | |||||||| |   | ||| | |
Db        362 CTGAGGTTGCACTCATCGTCTTCTCTAGCCGCGGTCGCCTTTATGAGTACTCTAACGATA 421

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||     || |||||||||||||| ||| | |  | ||          ||  |||
Db        422 GTGTCAAATCAACAATTGAGAGGTACAAAAAGGCATCTGCAGATTCTTCAAACACTGGGT 481

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  ||||| ||||||| |  || || || ||||  || ||  |    || |
Db        482 CTGTTTCTGAAGCCAATGCTCAGTACTACCAGCAAGAAGCTGCCAAGCTGCGTTCCCAAA 541
```

```
Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | ||||| ||| |||| |||||| || ||||||| | |||||  | ||||
Db        542 TTGGTAATTTGCAGAATTCAAACAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTG 601

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |||||||| |||    | ||  | |||||||| |||||||| ||| || | || ||||
Db        602 TGAAGGAACTTAAGAGTTTGGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCA 661

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| ||||| ||| ||| |||||     |||||||  |||||           |
Db        662 AAAAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGC 721

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCA---GAATAGCCACAACA 538
              | |  | ||  || |||| ||     |||||  | |||  | |    ||  | || | ||
Db        722 ACAACAATAACCAGCTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCA 781

Qy        539 CGAACATGTTATCAGCTCCGGAATATGATGCACT---------GCCCGCATTCGACTCTC 589
              |||  || |  |||       | | |   ||  |          |   |||  |||||||
Db        782 TGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTC 841

Qy        590 GAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAA 649
              | || |     || |  |||  | |  |  | ||   ||| ||  |||||| || || |
Db        842 GGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGA 901

Qy        650 CAACGCTTCAGCTTGGCTGA 669
                | |||||| | |  | |
Db        902 TGGCCCTTCAGTTAGTTTAA 921


RESULT 2
US-09-410-464-11
; Sequence 11, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
;  APPLICANT: Strauss et al.
;  TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
;  TITLE OF INVENTION: poplar and other plant species.
;  FILE REFERENCE: 53375
;  CURRENT APPLICATION NUMBER: US/09/410,464
;  CURRENT FILING DATE: 1999-10-01
;  EARLIER APPLICATION NUMBER: 09/287,700
;  EARLIER FILING DATE: 1999-04-06
;  EARLIER APPLICATION NUMBER: 60/080,851
;  EARLIER FILING DATE: 1998-04-06
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 723
;   TYPE: DNA
;   ORGANISM: Populus balsamifera subsp. trichocarpa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(723)
US-09-410-464-11

  Query Match             28.6%;  Score 259.8;  DB 4;  Length 723;
  Best Local Similarity   63.9%;  Pred. No. 5.1e-67;
  Matches  431;  Conservative    0;  Mismatches  232;  Indels   12;  Gaps    2;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db         47 TGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 106

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || | ||||||  | ||||| || || |||||||| |||||||| ||||
Db        107 TCTGCAAAAGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATG 166

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || || ||||||||||||||| ||| | ||  | |||||||| |    | ||| | |
Db        167 CTGAGGTTGCACTCATCGTCTTCTCTAGCCGCGGTCGCCTTTATGAGTACTCTAACGATA 226

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||    || |||||||||||||| ||| | |  |  ||        ||   |||

Db        227 GTGTCAAATCAACAATTGAGAGGTACAAAAAGGCATCTGCAGATTCTTCAAACACTGGGT 286

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  |||| ||||||| |  || || || |||  ||  ||  |    || |
Db        287 CTGTTTCTGAAGCCAATGCTCAGTACTACCAGCAAGAAGCTGCCAAGCTGCGTTCCCAAA 346

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | ||||| ||| |||| |||||| || ||||||| | |||||    | ||||
Db        347 TTGGTAATTTGCAGAATTCAAACAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTG 406

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |||||||| |||    | ||  | |||||||| |||||||| ||| || | || ||||
Db        407 TGAAGGAACTTAAGAGTTTGGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCA 466

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| ||||| ||| ||| |||||     |||||||  |||||           |
Db        467 AAAAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGC 526

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCA---GAATAGCCACAACA 538
              | |  | ||  || |||| ||     |||||  | |||  | |    ||  | || | ||
Db        527 ACAACAATAACCAGCTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCA 586

Qy        539 CGAACATGTTATCAGCTCCGGAATATGATGCACT---------GCCCGCATTCGACTCTC 589
              |||  || |  |||       | | |   ||  |          |   |||  |||||||
Db        587 TGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTC 646

Qy        590 GAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAA 649
              | || |     || |  |||  | |  |  | ||   ||| ||  |||||| || || |
Db        647 GGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGA 706

Qy        650 CAACGCTTCAGCTTG 664
                | |||||| | |
Db        707 TGGCCCTTCAGTTAG 721


RESULT 3
US-09-067-800-7
; Sequence 7, Application US/09067800
; Patent No. 6198024
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    APPLICANT:  Ferrandiz, Cristina
;    TITLE OF INVENTION:  Seed Plants Characterized by Delayed
;    TITLE OF INVENTION:  Seed Dispersal
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/067,800
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 2948
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  959 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  78..818
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..959
;       OTHER INFORMATION:  /note= "AGL5 cDNA and deduced
;       OTHER INFORMATION:  protein sequences."
US-09-067-800-7

  Query Match             28.1%;  Score 255.6;  DB 3;  Length 959;
  Best Local Similarity   68.3%;  Pred. No. 9.9e-66;
  Matches  354;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              || ||  | ||||||| |||||||||||||| || || |||||||| || || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || ||||||||| | ||||| || ||||  | || ||  | |||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || || ||  || |||||||||    ||||  | || || || |  |||||| ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || ||||   | || || || |||||||||||  |||||  |||  |   ||  |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||  | ||| ||||||| |  || || |||||   ||||||| |  ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||       | || |||   || || ||  |  | |||||     |||    | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | |||||   | |||||  ||||||||||| || || ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| |||||||||||  ||| |||||| ||   |||||| | |||||||  |   ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                | || ||   | ||| |||  |||||  | ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640


RESULT 4
US-09-349-677-7
; Sequence 7, Application US/09349677
; Patent No. 6288305
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    APPLICANT:  Ferrandiz, Cristina
;    TITLE OF INVENTION:  Seed Plants Characterized by Delayed
;    TITLE OF INVENTION:  Seed Dispersal
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/349,677
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/067,800
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 2948
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  959 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  78..818
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..959
;      OTHER INFORMATION:  /note= "AGL5 cDNA and deduced
;      OTHER INFORMATION:  protein sequences."
US-09-349-677-7

  Query Match             28.1%;  Score 255.6;  DB 3;  Length 959;
  Best Local Similarity   68.3%;  Pred. No. 9.9e-66;
  Matches  354;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              || ||  | ||||||| |||||||||||||| || || |||||||| || || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || ||||||||| | ||||| || ||||  | || ||  | |||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || || ||  || |||||||||    ||||  | || || || |  |||||| ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || ||||   | || || || |||||||||||  |||||  |||  |   ||  |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||  | ||| ||||||| |  || || |||||   ||||||| |  ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||       | || |||   || || ||  |  | |||||     |||    | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | |||||   | |||||  ||||||||||| || || ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| |||||||||||  ||| |||||| ||   |||||| | |||||||  |   ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                | || ||   | ||| |||  |||||  | ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640
```

```
RESULT 5
US-08-460-512-6
; Sequence 6, Application US/08460512
; Patent No. 5744693
;  GENERAL INFORMATION:
;    APPLICANT:  MEYEROWITZ, Elliot M.
;    APPLICANT:  YANOFSKY, Martin F.
;    APPLICANT:  MA, Hong
;    TITLE OF INVENTION:  PLANTS HAVING ALTERED FLORAL DEVELOPMENT
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard F. Trecartin
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,512
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/293,278
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/956,694
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-57322/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1125 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  110..853
US-08-460-512-6

  Query Match             28.0%;  Score 254.6;  DB 1;  Length 1125;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-65;
  Matches  356;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              |||| | || |||||||||||| || |||| ||||| || ||||| || || |||||||
Db        159 TGGGAAGAGGAAAGATTGAGATCAAACGGATCGAAAACACAACGAATCGTCAAGTCACTT 218

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||| | || ||||||||| | || ||||| |||||||||||| || || |||||||
Db        219 TCTGCAAGAGACGCAATGGTTTACTCAAAAAGGCCTATGAATTATCTGTGCTCTGTGATG 278

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || ||  | || |||||||| |||||||| ||||| ||||| | |||||| |||
Db        279 CTGAGGTTGCTTTGATTGTCTTCTCAAGCAGAGGCAGACTCTATGAGTATGCCAACAACA 338

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | |||||    || ||||||||||||||||||  ||||    ||   |    ||  ||
Db        339 GTGTGAAAGCAACAATTGAGAGGTACAAGAAAGCTTGTTCAGATTCCTCAAACACTGGTT 398

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | || || ||| ||||| |||||||||  || || || ||    |||||  |   ||| |

Db        399 CAATTTCCGAGGCCAATGCTCAGTATTATCAGCAAGAAGCCTCCAAACTGCGCGCACAAA 458

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | | |||| |||    || ||| |  || ||||||||     ||  | ||  ||| |
Db        459 TTGGAAATCTGCAGAATCAGAACAGGAACATGTTGGGTGAATCACTGGCTGCACTGAGCC 518

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |  || || ||| | || ||       | ||||||||||| || ||| | |  ||||||
Db        519 TCAGAGATCTGAAGAATCTGGAACAAAAAATTGAAAAAGGCATTAGCAAAATCAGATCCA 578

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| || || ||| || || ||    |||||||| ||||||        |
Db        579 AAAAGAATGAGCTGCTGTTTGCTGAAATTGAGTACATGCAGAAGAGGGAAATTGATTTAC 638

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526
              | |  | ||| ||   ||  |    ||||| || ||   | |||
Db        639 ACAACAACAATCAGTACCTGAGAGCAAAGATTGCTGAAACTGAGA 683


RESULT 6
US-09-410-464-15
; Sequence 15, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
;  APPLICANT: Strauss et al.
;  TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
;  TITLE OF INVENTION: poplar and other plant species.
;  FILE REFERENCE: 53375
;  CURRENT APPLICATION NUMBER: US/09/410,464
;  CURRENT FILING DATE: 1999-10-01
;  EARLIER APPLICATION NUMBER: 09/287,700
;  EARLIER FILING DATE: 1999-04-06
;  EARLIER APPLICATION NUMBER: 60/080,851
;  EARLIER FILING DATE: 1998-04-06
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 714
;   TYPE: DNA
;   ORGANISM: Populus balsamifera subsp. trichocarpa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(714)
US-09-410-464-15

  Query Match             27.3%;  Score 248.2;  DB 4;  Length 714;
  Best Local Similarity   67.0%;  Pred. No. 1.3e-63;
  Matches  352;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              |||| | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db         47 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 106

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || |||||||| | ||||| || |||||||||||| |||||||| ||||
Db        107 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 166

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || || |||||||||||||||||| | ||  | |||||||| |    | ||| | |
Db        167 CTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAATA 226

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||    || ||||| |||||||| ||| | || |  ||       | ||   |||
Db        227 GTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGGGT 286

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  ||||| |||||  |  || || || ||||  || ||  |    || |
Db        287 CAGTTTCTGAAGCCAATGCTCAGTTCTATCAGCAAGAAGCTGCCAAGCTGCGCTCGCAAA 346

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | |||||| ||| |||| ||| || || |||||||    |||| || ||||
Db        347 TTGGTAATTTGCAGAATTCAAACAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTG 406
```

```
Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
Db        407 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 466

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
Db        467 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 526

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526
Db        527 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 571
```

RESULT 7
US-09-410-464-14
; Sequence 14, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 53375
; CURRENT APPLICATION NUMBER: US/09/410,464
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 09/287,700
; EARLIER FILING DATE: 1999-04-06
; EARLIER APPLICATION NUMBER: 60/080,851
; EARLIER FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (99)..(815)
US-09-410-464-14

Query Match 27.3%; Score 248.2; DB 4; Length 1159;
Best Local Similarity 67.0%; Pred. No. 1.6e-63;
Matches 352; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
Db        145 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 204

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
Db        205 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 264

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
Db        265 CTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAATA 324

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
Db        325 GTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGGGT 384

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
Db        385 CAGTTTCTGAAGCCAATGCTCAGTTCTATCAGCAAGAAGCTGCCAAGCTGCGCTCGCAAA 444

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
Db        445 TTGGTAATTTGCAGAATTCAAACAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTG 504

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
Db        505 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 564

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
```

```
Db        565 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 624

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526
Db        625 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 669
```

RESULT 8
US-09-433-579-1
; Sequence 1, Application US/09433579
; Patent No. 6444877
; GENERAL INFORMATION:
; APPLICANT: Rottmann, William H.
; TITLE OF INVENTION: LSAG Gene
; FILE REFERENCE: LSAG Gene
; CURRENT APPLICATION NUMBER: US/09/433,579
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Liquidambar styraciflua
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (331)..(1008)
US-09-433-579-1

Query Match 27.3%; Score 248.2; DB 4; Length 1297;
Best Local Similarity 62.6%; Pred. No. 1.7e-63;
Matches 426; Conservative 0; Mismatches 243; Indels 12; Gaps 2;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db        331 ATGGGTAGGGGAAAGATCGAGATCAAGCGGATCGAGAACACGACAAACCGCCAAGTCACC 390

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db        391 TTCTGTAAACGGCGCAACGGCTTACTCAAGAAGGCCTATGAATTATCTGTTCTCTGTGAT 450

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        451 GCCGAGGTAGCCCTTATCGTCTTCTCTACCCGTGGCCGCCTTTACGAGTATGCCAACAAC 510

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        511 AGTGTTAAAGCAACAATTGAGAGGTATAAGAAGGCATCTGTGGATTCCTCCAATACTGGA 570

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        571 TCTGTTTCTGAAGCCAATGCTCAGTTCTACCAGCAAGAAGCTGGCAAACTGCGTAACCAA 630

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
Db        631 ATCAGGAATATGCAGAATACAAACAGGACTATGCTGGGTGAGTCTTTGGGCTCTCTGAGT 690

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
Db        691 CCCAAAGAACTCAAGGGTCTGGAGACTAAATTAGAGAAAGGCATTAGCAAAATAAGGTCC 750

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
Db        751 AAAAAGAATGAGCTACTGTTTTCCGAAATTGAATACATGCAAAAGAGGGAAATCGACTTG 810

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA---ATAGCCACAAC 537
Db        811 CACAATGATAACCAGTATCTCCGAGCAAAAATAGCTGAACACGAGAGAGCTCAACAGCAA 870

Qy        538 ACGAACATGTTATCAGCTCCGGAATATGATG---------CACTGCCCGCATTCGACTCT 588
Db        871 ATGAACTTGATGCCAGGTGGATCAAACTATGAGGTCTTGCCATCACAGCCATTTGACTCT 930

Qy        589 CGAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAA 648
```

```
Db        931 CGGAACTTTTTCCAAGTAAATGCATTGCAACCCAATCATCATTACTCTCGCCAAGATCAA 990

Qy        649 ACAACGCTTCAGCTTGGCTGA 669
              |    |||||   |  |   |
Db        991 ATGGCCCTTCAATTAGTTTAA 1011


RESULT 9
US-09-067-800-5
; Sequence 5, Application US/09067800
; Patent No. 6198024
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    APPLICANT:  Ferrandiz, Cristina
;    TITLE OF INVENTION:  Seed Plants Characterized by Delayed
;    TITLE OF INVENTION:  Seed Dispersal
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/067,800
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 2948
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  896 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7..753
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  896
;      OTHER INFORMATION:  /note= "There is a poly(A) tail at
;      OTHER INFORMATION:  the end of the cDNA sequence."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..896
;      OTHER INFORMATION:  /note= "AGL1 cDNA and deduced
;      OTHER INFORMATION:  protein sequences."
US-09-067-800-5

  Query Match             27.0%;  Score 245;  DB 3;  Length 896;
  Best Local Similarity   67.1%;  Pred. No. 1.3e-62;
  Matches  347;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              | ||  |  |||||  || |||||||||||||  || || || || || || || ||||
Db         53 TAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACAAATCGTCAAGTTACTT 112

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  ||  || ||||||  |  | |||||  ||  ||||||  |  ||  ||  | |||||||

Db        113 TCTGCAAACGACGCAATGGTCTTCTCAAGAAAGCTTATGAACTCTCTGTCTTGTGTGATG 172

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | |||||  ||||||  ||  |||||||||    |  ||  |  ||  |||||  |  ||||||  |||
Db        173 CCGAAGTTGCCCTCGTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAACA 232

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | ||||  |  |  ||  |||||  ||||||||||  ||||    ||    |  |  ||
Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              |  |  | ||  | |||  ||||||||  |  ||  ||  ||  ||  |||  ||  |  ||||
Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              ||     |  |  ||  |||  ||||||||||  |  |  ||  ||    |||    |  ||||||
Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              |  ||||||||||  |  ||  ||  |  ||  ||||||||||||||||||||  ||  ||  ||||
Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              |||||||  |||  ||||  |  |  ||||||  ||  ||||||  ||||||  |
Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                ||  |  ||   |   ||  ||   ||||||||||
Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569


RESULT 10
US-09-349-677-5
; Sequence 5, Application US/09349677
; Patent No. 6288305
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    APPLICANT:  Ferrandiz, Cristina
;    TITLE OF INVENTION:  Seed Plants Characterized by Delayed
;    TITLE OF INVENTION:  Seed Dispersal
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/349,677
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/067,800
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 2948
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  896 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  7..753
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  896
;      OTHER INFORMATION:  /note= "There is a poly(A) tail at
;      OTHER INFORMATION:  the end of the cDNA sequence."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..896
;      OTHER INFORMATION:  /note= "AGL1 cDNA and deduced
;      OTHER INFORMATION:  protein sequences."
US-09-349-677-5

  Query Match             27.0%;  Score 245;  DB 3;  Length 896;
  Best Local Similarity   67.1%;  Pred. No. 1.3e-62;
  Matches  347;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              |  |   | |||||  || ||||||||||||||| || ||  ||  ||  ||  ||  |||| 
Db         53 TAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACAAATCGTCAAGTTACTT 112

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  ||  ||  ||||||  |   | |||||  ||  ||||||  |  ||  ||  |||||||
Db        113 TCTGCAAACGACGCAATGGTCTTCTCAAGAAAGCTTATGAACTCTCTGTCTTGTGTGATG 172

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              |  |||||  ||||||  ||  |||||||||   |  ||   |  ||  |||||  |   ||||||  |||
Db        173 CCGAAGTTGCCCTCGTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAACA 232

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              |  ||||  |  |  ||  |||||  ||||||||||   ||||     ||    |  |  ||
Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              |   |  | ||  | |||  |||||||  |  ||  ||  ||  ||    |||  ||   |     ||||
Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              ||      |  |  ||  |||  ||||||||||  |   |  ||  ||     |||     |  ||||||
Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              |  ||||||||||   |  ||  ||  |  ||  ||||||||||||||||||||  ||  ||  ||||
Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              |||||||  |||  ||||   |  |  ||||||  ||     |||||||  ||||||          |
Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                ||  |||   |   ||  ||    ||||||||||
Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569


RESULT 11
US-08-460-512-1
; Sequence 1, Application US/08460512
; Patent No. 5744693
;  GENERAL INFORMATION:
;    APPLICANT:  MEYEROWITZ, Elliot M.
;    APPLICANT:  YANOFSKY, Martin F.
;    APPLICANT:  MA, Hong
;    TITLE OF INVENTION:  PLANTS HAVING ALTERED FLORAL DEVELOPMENT
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard F. Trecartin
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,512
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/293,278
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/956,694
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-57322/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1457 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  416..1270
US-08-460-512-1

  Query Match             25.6%;  Score 232.4;  DB 1;  Length 1457;
  Best Local Similarity   64.9%;  Pred. No. 8.2e-59;
  Matches  344;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy          4 GGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTTTC 63
              ||  | ||  |||||  ||  ||  ||    ||||  ||  ||  ||  |||||  ||  ||  ||||||||
Db        566 GGGAGAGGAAAGATCGAAATCAAACGGATCGAGAACACAACGAATCGTCAAGTCACTTTT 625

Qy         64 TGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATGCA 123
              |||||  ||  ||||||||||  |  |||||  ||  ||  ||  |  ||  |||||  ||||||||
Db        626 TGCAAACGTAGAAATGGTTTGCTCAAGAAAGCTTACGAGCTCTCTGTTCTCTGTGATGCT 685

Qy        124 GAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACAGC 183
              |||||  ||  ||||||||  |||||  |||  |  ||  |  ||  ||||  |   |  |||  ||||
Db        686 GAAGTCGCACTCATCGTTTTCTCTAGCCGTGGTCGTCTCTATGAGTACTCTAACAACAGT 745

Qy        184 GTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGGCG 243
              ||  ||  |  ||  ||||||||||||||||||||  |       |||||      ||   ||  ||
Db        746 GTAAAAGGTACTATTGAGAGGTACAAGAAGGCAATATCGGACAATTCTAACACCGGATCG 805

Qy        244 ATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGATT 303
              |  ||||    |||  |  |||||||  |||||  ||  |  |  |||  |  |  |||||  |||
Db        806 GTGGCAGAAATTAATGCACAGTATTATCAACAAGAATCAGCCAAATTGCGTCAACAAATT 865

Qy        304 GACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACATT 363
              ||    |  |||||  |  ||  |||||  ||||||||||    |  |     |   ||
Db        866 ATCAGCATACAAAACTCCAACAGGCAATTGATGGGTGAGACGATAGGGTCAATGTCTCCC 925

Qy        364 AAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCAAA 423
              ||  ||  ||||  |  |  |  ||  |  ||  |  |  ||  |  ||||  |  ||||||||
Db        926 AAAGAGCTCAGGAACTTGGAAGGCAGATTAGAGAGAAGTATTACCCGAATCCGATCCAAG 985

Qy        424 AAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTATC 483
              |||||  |||  |  ||   ||  ||||||  ||||||||  |||  |||       |
Db        986 AAGAATGAGCTCTTATTTTCTGAAATCGACTACATGCAGAAAAGAGAAGTTGATTTGCAT 1045

Qy        484 CAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCA 533
              |  ||  ||  ||||||||||     ||||||||  ||    |  |||  |  |  |
Db       1046 AACGATAACCAGATTCTTCGTGCAAAGATAGCTGAAAATGAGAGGAACAA 1095
```

```
RESULT 12
US-08-460-512-3
; Sequence 3, Application US/08460512
; Patent No. 5744693
;  GENERAL INFORMATION:
;    APPLICANT:  MEYEROWITZ, Elliot M.
;    APPLICANT:  YANOFSKY, Martin F.
;    APPLICANT:  MA, Hong
;    TITLE OF INVENTION:  PLANTS HAVING ALTERED FLORAL DEVELOPMENT
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard F. Trecartin
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,512
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/293,278
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/956,694
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-57322/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1097 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..884
US-08-460-512-3

  Query Match             25.4%;  Score 230.8;  DB 1;  Length 1097;
  Best Local Similarity   64.7%;  Pred. No. 2.1e-58;
  Matches  343;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy          4 GGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTTTC 63
              ||  | || ||||| || |||||  |||| || || || |||||||| || || ||||||
Db        177 GGGAGAGGAAAGATCGAAATAAAACGGATCGAGAACACAACGAACCGTCAAGTTACTTTC 236

Qy         64 TGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATGCA 123
              ||||| ||| ||||||||||   |||||  || || ||| | || ||||||||||||||
Db        237 TGCAAACGCAGAAATGGTTTGCTCAAGAAAGCTTACGAACTCTCTGTTCTTTGTGATGCT 296

Qy        124 GAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACAGC 183
              ||||| || |||||  |||||||| ||| | ||  | || |||||  |    | ||| ||||
Db        297 GAAGTCGCACTCATTGTCTTCTCTAGCCGTGGCCGTCTCTATGAGTACTCAAACAACAGT 356

Qy        184 GTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGGCG 243
              || ||  |||| |||||||||||||||||  |        || ||      ||  || |
Db        357 GTAAAAGGGACAATTGAGAGGTACAAGAAAGCAATATCGGATAATTCTAACACCGGATCC 416

Qy        244 ATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGATT 303
              |  ||||    ||| | |||||||  ||||| ||  ||| ||| |  | ||||| |||
Db        417 GTGGCAGAAATTAATGCACAGTATTATCAACAAGAATCTGCCAAATTGCGTCAACAAATT 476

Qy        304 GACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACATT 363
              ||    | || ||  | ||| ||||| ||||||||||    | |||      | ||
Db        477 ATCAGCATACAGAACTCGAACAGGCAATTGATGGGTGAGACGATTGGGTCAATGTCTCCC 536

Qy        364 AAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCAAA 423
              || || |||| | |  | || |   || | || | | | | | ||| | ||||||||
Db        537 AAAGAGCTCAGGAACTTGGAAGGCAGATTAGACAGAAGTGTTAATCGAATCCGATCCAAG 596

Qy        424 AAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTATC 483
              ||||||||  | ||  | |  || || |||  |||||||  || |||          |
Db        597 AAGAACGAACTCTTATTCGCCGAAATTGACTACATGCAGAAGAGAAGAAGTTGATTTGCAT 656

Qy        484 CAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCA 533
              | || ||  || |||||||    ||||||||  ||   | |||  | | |
Db        657 AACGATAACCAGCTTCTTCGTGCTAAGATAGCTGAAAATGAGAGGAACAA 706


RESULT 13
US-08-867-087B-12
; Sequence 12, Application US/08867087B
; Patent No. 5990386
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gynheung
;    TITLE OF INVENTION:  GENES CONTROLLING FLORAL DEVELOPMENT
;    TITLE OF INVENTION:  AND APICAL DOMINANCE IN PLANTS
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klarquist Sparkman Campbell Leigh &
;      ADDRESSEE:  Whinston, LLP
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  United States of America
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 3-1/2 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,087B
;      FILING DATE:  June 2, 1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 08/323,449
;      FILING DATE:  October 14, 1994
;      APPLICATION NUMBER:  U.S. 08/485,981
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Alan. E.
;      REGISTRATION NUMBER:  35,123
;      REFERENCE/DOCKET NUMBER:  4630-47071
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1043 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
US-08-867-087B-12

  Query Match             16.2%;  Score 147.2;  DB 2;  Length 1043;
  Best Local Similarity   58.5%;  Pred. No. 9.2e-34;
  Matches  276;  Conservative    0;  Mismatches  193;  Indels    3;  Gaps    1;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
```

```
Db         34 ATGGGGAGGGGAAGAGTTGAGCTGAAGCGCATCGAGAACAAGATCAACAGGCAGGTCACC 93

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db         94 TTCTCCAAGCGCCGCAACGGCCTCCTCAAGAAGGCCTACGAGCTGTCCGTTCTCTGCGAC 153

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        154 GCCGAGGTCGCGCTCATCATCTTCTCCAGCCGCGGCAAGCTCTACGAGTTCGGCAGCGCC 213

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        214 GGCATAACAAAGACTTTAGAAAGGTACCAACATTGTTGC---TACAATGCTCAAGATTCC 270

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        271 AACAATGCACTTTCTGAAACTCAGAGTTGGTACCATGAAATGTCAAAGTTGAAAGCAAAA 330

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        331 TTTGAAGCTTTGCAGCGCACTCAAAGGCACTTGCTTGGGGAGGATCTTGGACCACTCAGC 390

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        391 GTCAAAGAATTGCAGCAGCTGGAGAAACAGCTTGAATGTGCACTATCACAGGCGAGACAG 450

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAAC 472

Db        451 AGAAAGACGCAACTGATGATGGAACAGGTGGAGGAACTTCGCAGAAAGGAGC 502


RESULT 14
US-09-067-800-1
; Sequence 1, Application US/09067800
; Patent No. 6198024
;  GENERAL INFORMATION:
;    APPLICANT:  Yanofsky, Martin F.
;    APPLICANT:  Ferrandiz, Cristina
;    TITLE OF INVENTION:  Seed Plants Characterized by Delayed
;    TITLE OF INVENTION:  Seed Dispersal
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/067,800
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 2948
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1062 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..827
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1062
;      OTHER INFORMATION:  /note= "There is a poly(A) tail at
;      OTHER INFORMATION:  the end."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..1062
;      OTHER INFORMATION:  /note= "Nucleotide and Deduced
;      OTHER INFORMATION:  Amino Acid Sequences of the AGL8 cDNA clone."
US-09-067-800-1

  Query Match             15.6%;  Score 141.6;  DB 3;  Length 1062;
  Best Local Similarity   55.5%;  Pred. No. 4.2e-32;
  Matches  295;  Conservative    0;  Mismatches  234;  Indels    3;  Gaps    1;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db        101 ATGGGAAGAGGTAGGGTTCAGCTGAAGAGGATAGAGAACAAGATCAATAGGCAAGTTACT 160

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        161 TTCTCAAAGAGAAGGTCTGGTTTGCTCAAGAAAGCTCATGAGATCTCTGTTCTCTGCGAT 220

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        221 GCTGAGGTTGCTCTCATCGTCTTCTCTTCCAAAGGCAAACTCTTCGAATATTCCACCGAC 280

Qy        181 ---AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGA 237

Db        281 TCTTGCATGGAGAGGATACTTGAACGCTATGATCGCTATTTATATTCAGACAAACAACTT 340

Qy        238 GGGGCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAA 297

Db        341 GTTGGCCGAGACGTTTCACAAAGTGAAAATTGGGTTCTAGAACATGCTAAGCTCAAGGCA 400

Qy        298 CAGATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTG 357

Db        401 AGAGTTGAGGTACTTGAGAAGAACAAAAGGAATTTTATGGGGGAAGATCTTGATTCGTTG 460

Qy        358 AACATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGA 417

Db        461 AGCTTGAAGGAGCTCCAAAGCTTGGAGCATCAGCTCGATGCAGCTATCAAGAGCATTAGG 520

Qy        418 TCCAAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATA 477

Db        521 TCAAGAAAGAACCAAGCTATGTTCGAATCCATATCTGCGCTCCAGAAGAAGGATAAAGCC 580

Qy        478 CTTATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATA 529

Db        581 TTGCAAGATCACAACAATTCGCTTCTCAAAAAGATTAAGGAGAGGGAGAAGA 632


RESULT 15
US-09-105-652-1
; Sequence 1, Application US/09105652
; Patent No. 6229068
; GENERAL INFORMATION:
;  APPLICANT: Yanofsky, Martin F.
;  APPLICANT: Martienssen, Robert
;  APPLICANT: Ferrandiz, Cristina
;  APPLICANT: Gu, Qing
;  TITLE OF INVENTION: Method of Increasing Fruit Size in a Plant
;  FILE REFERENCE: P-UD 3040
;  CURRENT APPLICATION NUMBER: US/09/105,652
;  CURRENT FILING DATE: 1998-06-26
;  EARLIER APPLICATION NUMBER: 60/051,030
;  EARLIER FILING DATE: 1997-06-27
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
```

; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(826)
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: (1061)
US-09-105-652-1

Query Match 15.6%; Score 141.6; DB 3; Length 1062;
Best Local Similarity 55.5%; Pred. No. 4.2e-32;
Matches 295; Conservative 0; Mismatches 234; Indels 3; Gaps 1;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || | | || || | |||||||| || || |  |  ||  | || || |||
Db        101 ATGGGAAGAGGTAGGGTTCAGCTGAAGAGGATAGAGAACAAGATCAATAGGCAAGTTACT 160

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||  ||| |  |   ||||||  | ||||| ||  ||||  | || |||||  || |||
Db        161 TTCTCAAAGAGAAGGTCTGGTTTGCTCAAGAAAGCTCATGAGATCTCTGTTCTCTGCGAT 220

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || || ||||||||||||||    || ||| | ||| |  |||| | ||| | ||
Db        221 GCTGAGGTTGCTCTCATCGTCTTCTCTTCCAAAGGCAAACTCTTCGAATATTCCACCGAC 280

Qy        181 ---AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGA 237
                 || || ||||||   ||||  | ||  |       ||     |    |||| ||
Db        281 TCTTGCATGGAGAGGATACTTGAACGCTATGATCGCTATTTATATTCAGACAAACAACTT 340

Qy        238 GGGGCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAA 297
              |  |   |   |  ||  |  | |  |||||  |  ||  || ||| ||||     |
Db        341 GTTGGCCGAGACGTTTCACAAAGTGAAAATTGGGTTCTAGAACATGCTAAGCTCAAGGCA 400

Qy        298 CAGATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTG 357
                ||||  |  |  | ||    || ||| |||| ||||| || |  |||    | |||
Db        401 AGAGTTGAGGTACTTGAGAAGAACAAAAGGAATTTTATGGGGGAAGATCTTGATTCGTTG 460

Qy        358 AACATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGA 417
              | | | ||||| ||| |      | |||  ||  || ||  ||  ||||  |  |  |
Db        461 AGCTTGAAGGAGCTCCAAAGCTTGGAGCATCAGCTCGATGCAGCTATCAAGAGCATTAGG 520

Qy        418 TCCAAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATA 477
              || | |||||||  |    || | |||   ||        | ||||  | |||  |
Db        521 TCAAGAAAGAACCAAGCTATGTTCGAATCCATATCTGCGCTCCAGAAGAAGGATAAAGCC 580

Qy        478 CTTATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATA 529
               |     |  | ||  |   |||| ||  |||||    ||| |  |||| |
Db        581 TTGCAAGATCACAACAATTCGCTTCTCAAAAAGATTAAGGAGAGGGAGAAGA 632
```

Search completed: January 21, 2004, 22:19:47
Job time : 64.4877 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 3315.05 Seconds
(without alignments)
11217.584 Million cell updates/sec

Title: US-09-936-869-3
Perfect score: 909
Sequence: 1 atgggtcgtgggaagattga.....cgcatgttatctttctgttg 909

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	909	100.0	922	8	AF023615	AF023615 Pinus rad
2	885	97.4	1726	8	AF006210	AF006210 Pinus res
3	782.8	86.1	1555	8	PMU69483	U69483 Picea maria
4	782.8	86.1	1642	8	PMU46582	U46582 Picea maria
5	778	85.6	1657	8	PMU69482	U69482 Picea maria
6	774.8	85.2	1753	8	PADAL2	X79280 P.abies dal
7	772.8	85.0	1212	8	PMU69484	U69484 Picea maria
8	455.6	50.1	1372	8	AY114304	AY114304 Ginkgo bi
9	433	47.6	1164	8	AF492455	AF492455 Cycas ede
10	328.6	36.1	1154	8	GGN132209	AJ132209 Gnetum gn
11	303.8	33.4	912	8	AB050644	AB050644 Magnolia
12	279.6	30.8	942	8	AF373604	AF373604 Vitis vin
13	279	30.7	1009	8	AF234617	AF234617 Phalaenop
14	276.8	30.5	1142	8	AF022377	AF022377 Cucumis s
15	266	29.3	1001	8	AY178837	AY178837 Momordica
16	265.8	29.2	1160	8	AF265562	AF265562 Vitis vin
17	258.2	28.4	693	8	AF538966	AF538966 Gossypium
18	256.4	28.2	1048	8	AF035439	AF035439 Cucumis s
19	256.2	28.2	1131	8	SLSLM1	X80488 S.latifolia
20	255.6	28.1	959	6	AR138348	AR138348 Sequence
21	255.6	28.1	963	8	ATHAGL5A	M55553 Arabidopsis
22	255	28.1	869	8	AB079260	AB079260 Agapanthu
23	254.6	28.0	1065	8	MDO251117	AJ251117 Malus dom
24	254.6	28.0	1125	6	AR003863	AR003863 Sequence
25	254.6	28.0	1125	8	TOBNAG1A	L23925 Nicotiana t
26	254	27.9	1165	8	BPE252071	AJ252071 Betula pe
27	253	27.8	1084	8	TOMTAG1A	L26295 Lycopersico
28	251.4	27.7	1157	6	AX006357	AX006357 Sequence
29	251.4	27.7	1158	8	AF035438	AF035438 Cucumis s
30	251.4	27.7	1160	8	AF022379	AF022379 Cucumis s
31	251	27.6	1214	8	PHPMADS3	X72912 P.hybrida p
32	250.4	27.5	1062	8	AF194335	AF194335 Hyacinthu
33	250.2	27.5	1061	8	AF099937	AF099937 Hyacinthu
34	248.2	27.3	1297	6	AR227777	AR227777 Sequence
35	247.2	27.2	1179	8	AF027376	AF027376 Corylus a
36	245.4	27.0	1080	8	PGORFGAG2	Z46612 P.ginseng g
37	245	27.0	895	8	ATHAGL1A	M55550 Arabidopsis
38	245	27.0	896	6	AR138347	AR138347 Sequence
39	245	27.0	1202	8	AY086196	AY086196 Arabidops
40	243.6	26.8	1037	8	MDO251118	AJ251118 Malus dom
41	242.2	26.6	967	8	CSCUS1	X97801 C.sativus C
42	241.8	26.6	1065	8	AB025643	AB025643 Rosa rugo
43	241.6	26.6	916	8	AY098736	AY098736 Lycopersi
44	241.2	26.5	1000	8	AF022378	AF022378 Cucumis s
45	240.4	26.4	1073	8	S53900	S53900 ple=plena [

ALIGNMENTS

RESULT 1
AF023615
LOCUS AF023615 922 bp mRNA linear PLN 26-JAN-1999
DEFINITION Pinus radiata MADS box protein mRNA, complete cds.
ACCESSION AF023615
VERSION AF023615.1 GI:4103485
KEYWORDS .
SOURCE Pinus radiata (Monterey pine)
ORGANISM Pinus radiata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 922)
AUTHORS Jun-Jun Liu and G. K. Podila.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1997) Department of Biological Sciences, Michigan

```
           Technological University, Houghton, MI 49931, USA
FEATURES             Location/Qualifiers
     source          1. .922
                     /organism="Pinus radiata"
                     /mol_type="mRNA"
                     /db_xref="taxon:3347"
     CDS             14. .682
                     /codon_start=1
                     /product="MADS box protein"
                     /protein_id="AAD09342.1"
                     /db_xref="GI:4103486"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQID
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGISRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      303 a    174 c    218 g    227 t
ORIGIN

  Query Match           100.0%;  Score 909;  DB 8;  Length 922;
  Best Local Similarity  100.0%;  Pred. No. 4.3e-238;
  Matches  909;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 73

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 133

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 193

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 253

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 313

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 373

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 433

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 493

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 553

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        554 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 613

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        614 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 673

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 733

Qy        721 CTAGTAACCTTGAATGAGATTCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        734 CTAGTAACCTTGAATGAGATTCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGA 793

Qy        781 GATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        794 GATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 853

Qy        841 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        854 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 913

Qy        901 TTTCTGTTG 909
              |||||||||
Db        914 TTTCTGTTG 922


RESULT 2
AF006210
LOCUS       AF006210                1726 bp    mRNA    linear   PLN 05-JAN-1999
DEFINITION  Pinus resinosa MADS box transcription factor mRNA, complete cds.
ACCESSION   AF006210
VERSION     AF006210.1  GI:4101709
KEYWORDS    .
SOURCE      Pinus resinosa
  ORGANISM  Pinus resinosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 1726)
  AUTHORS   Liu,J.-J. and Podila,G.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1997) Department of Biological Sciences, Michigan
            Technological University, Houghton, MI 49931, USA
FEATURES             Location/Qualifiers
     source          1. .1726
                     /organism="Pinus resinosa"
                     /mol_type="mRNA"
                     /db_xref="taxon:54921"
                     /sex="female"
                     /dev_stage="immature cones"
                     /note="Lambda TriplEx cDNA library"
     CDS             475. .1143
                     /codon_start=1
                     /product="MADS box transcription factor"
                     /protein_id="AAD01266.1"
                     /db_xref="GI:4101710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGISRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      527 a    346 c    382 g    471 t
ORIGIN

  Query Match           97.4%;  Score 885;  DB 8;  Length 1726;
  Best Local Similarity  98.3%;  Pred. No. 1.7e-231;
  Matches  894;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 534

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        535 TTCTGCAAGCGCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 594

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 654

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db        655 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCATGGGGGG 714

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
```

```
Db        715 GCGATTTCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 774

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        775 ATTGAAATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 834

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        835 ATCAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTACGATCC 894

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        895 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 954

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        955 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 1014

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db       1015 AACATGTTATCAGCTCCCGAATATGATGCATTGCCAGCATTCGACTCTCGAAATTTTCTA 1074

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Db       1075 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 1134

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720

Db       1135 CTTGGCTGAACGGTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 1194

Qy        721 CTAGTAACCTTGAATGAGATTCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGA 780

Db       1195 TTAGTAACCTTGAATGAGATTCAGAGTCCAAATATTGCGAGGCAAGAGCACAATGGAAGA 1254

Qy        781 GATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 840

Db       1255 GATAGCTCCTAGTATGAATATGGGTTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 1314

Qy        841 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 900

Db       1315 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 1374

Qy        901 TTTCTGTTG 909

Db       1375 TTTCTGTTG 1383
```

RESULT 3
PMU69483

```
LOCUS       PMU69483                1555 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SMADS42C
            mRNA, complete cds.
ACCESSION   U69483
VERSION     U69483.1  GI:4033722
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1555)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1555)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1555
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             329. .997
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SMADS42C"
                     /protein_id="AAC97158.1"
                     /db_xref="GI:4033723"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      476 a    297 c    348 g    434 t
ORIGIN
```

```
  Query Match             86.1%;  Score 782.8;  DB 8;  Length 1555;
  Best Local Similarity   93.1%;  Pred. No. 1.8e-203;
  Matches  855;  Conservative    0;  Mismatches   52;  Indels   11;  Gaps    3;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db        329 ATGGGCCGTGGGAAGATTGAGATAAAGAGGATTGAGAATACTACGAACCGACAGGTCACT 388

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        389 TTCTGTAAGCGCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 448

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        449 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAGTTTGCCAACCAC 508

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        509 AGCGTGAAGAGAACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 568

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        569 GCGATTTCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 628

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        629 ATTGAAATTTTGCAAAATGCAAATAGGCACTTGATGGGTGACGGGCTTACAGCTTTAAAC 688

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        689 ATCAAGGAACTCAAGCAGCTTGAGGTTCGACTTGAAAAAGGAATCGGCCGAGTACGATCC 748

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        749 AAGAAGAACGAGATGTTGCTTGAGGAGATCGACATCATGCAGAGAAGGGAACACATTCTT 808

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        809 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGCCAGAATAGCCACAACACG 868

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        869 AGCATGCTATCAGCTCCCGAATATGATGCATTGCCAGCATTCGATTCTCGAAATTTCCTA 928

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Db        929 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAGCAAACAACGCTTCAG 988

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720
```

```
Db        989 CTTGGCTGAAGGTTGAAGTGGTGGATGCTTAAAACTCAATCAAGGTCCTTGAAAAATATG 1048

Qy        721 CTAGTAACCTTGAATGAGAT---------TCAGAGTCGAAATATTGCGAGGCAAGAGCAC 771

Db       1049 -TTGTAGCCTTGAATGAGATTCCCATGGATCCCAGTCGAAATATTGTGAGGCAAAATCAC 1107

Qy        772 AATGGAAGAGATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCA 831

Db       1108 AATGAAGGAGATAGCTCCTAGTATAAATATGGATTTATGATATTAACCTATGGTTTGTCA 1167

Qy        832 GCTTTAAATATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCG 891

Db       1168 GTTTTATATATAGCTGTTTGAAACAAAGAATAGAACATATTAGCTA-TTGTTTTCTCGCG 1226

Qy        892 CATGTTATCTTTCTGTTG 909

Db       1227 CATTTTAGCTTTCTGTTG 1244


RESULT 4
PMU46582
LOCUS       PMU46582                1642 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcription factor SMADS42B
            mRNA, complete cds.
ACCESSION   U46582
VERSION     U46582.1  GI:4033709
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1642)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewardt,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1996) Robert G. Rutledge, Molecular Genetics and
            Tissue Culture, Petawawa National Forestry Institue, Chalk River,
            ON K0J 1J0, Canada
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             390. .1058
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcription factor
                     SMADS42B"
                     /protein_id="AAC97146.1"
                     /db_xref="GI:4033710"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTSMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      485 a    321 c    366 g    470 t
ORIGIN

  Query Match           86.1%;  Score 782.8;  DB 8;  Length 1642;
  Best Local Similarity  93.1%;  Pred. No. 1.8e-203;
  Matches  855;  Conservative    0;  Mismatches   52;  Indels   11;  Gaps    3;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db        390 ATGGGCCGTGGGAAGATTGAGATAAAGAGGATTGAGAATACTACGAACCGACAGGTCACT 449

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        450 TTCTGTAAGCGCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 509

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        510 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAGTTTGCCAACCAC 569

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        570 AGCGTGAAGAGAACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 629

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        630 GCGATTTCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 689

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        690 ATTGAAATTTTGCAAAATGCAAATAGGCACTTGATGGGTGACGGGCTTACAGCTTTAAAC 749

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        750 ATCAAGGAACTCAAGCAGCTTGAGGTTCGACTTGAAAAAGGAATCGGCCGAGTACGATCC 809

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        810 AAGAAGAACGAGATGTTGCTTGAGGAGATCGACATCATGCAGAGAAGGGAACACATTCTT 869

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        870 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGCCAGAATAGCCACAACACG 929

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        930 AGCATGCTATCAGCTCCCGAATATGATGCATTGCCAGCATTCGATTCTCGAAATTTCCTA 989

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Db        990 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAGCAAACAACGCTTCAG 1049

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720

Db       1050 CTTGGCTGAAGGTTGAAGTGGTGGATGCTTAAAACTCAATCAAGGTCCTTGAAAAATATG 1109

Qy        721 CTAGTAACCTTGAATGAGAT---------TCAGAGTCGAAATATTGCGAGGCAAGAGCAC 771

Db       1110 -TTGTAGCCTTGAATGAGATTCCCATGGATCCCAGTCGAAATATTGTGAGGCAAAATCAC 1168

Qy        772 AATGGAAGAGATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCA 831

Db       1169 AATGAAGGAGATAGCTCCTAGTATAAATATGGATTTATGATATTAACCTATGGTTTGTCA 1228

Qy        832 GCTTTAAATATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCG 891

Db       1229 GTTTTATATATAGCTGTTTGAAACAAAGAATAGAACATATTAGCTA-TTGTTTTCTCGCG 1287

Qy        892 CATGTTATCTTTCTGTTG 909

Db       1288 CATTTTAGCTTTCTGTTG 1305


RESULT 5
PMU69482
LOCUS       PMU69482                1657 bp    mRNA    linear   PLN 24-JAN-2000
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SAG1a
            mRNA, complete cds.
ACCESSION   U69482
VERSION     U69482.1  GI:4033720
```

```
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1657)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1657)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1657
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             457. .1125
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SAG1a"
                     /protein_id="AAC97157.1"
                     /db_xref="GI:4033721"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
BASE COUNT      497 a    336 c    362 g    462 t
ORIGIN
```

```
  Query Match             85.6%;  Score 778;  DB 8;  Length 1657;
  Best Local Similarity   92.8%;  Pred. No. 3.7e-202;
  Matches  852;  Conservative    0;  Mismatches   55;  Indels   11;  Gaps    3;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||||| |||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        457 ATGGGCCGTGGGAAGATTGAGATAAAGAGGATTGAGAATACTACGAACCGACAGGTCACT 516

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        517 TTCTGTAAGCGCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 576

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        577 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAGTTTGCCAACCAC 636

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 AGCGTGAAGAGAACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 696

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        697 GCGATTTCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAGCAG 756

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||||| ||||| |||||||||||||||| |||||||||| |||||||||||||||| |||
Db        757 ATTGAAATTTTACAAAATGCAAATAGGCACTTGATGGGTGATGGGCTTACAGCTTTAAAC 816

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              || |||||||||||||| |||||||||||||||||||||||||| ||||||| ||||||
Db        817 ATCAAGGAACTCAAGCAGCTTGAGGTTCGACTTGAAAAAGGAATCGGCCGAGTACGATCC 876

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || |||||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db        877 AAGAAGAACGAGATGTTGCTTGAGGAGATCGACATCATGCAGAGAAGGGAACACATTCTT 936

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        937 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGCCAGAATAGCCACAACACG 996

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              |||||| |||||||||| |||||||||||| |||| |||||||| ||||||||||||||||
Db        997 AACATGCTATCAGCTCCCGAATATGATGCATTGCCAGCATTCGATTCTCGAAATTTCCTA 1056

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       1057 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAGCAAACAACGCTTCAG 1116

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720
              |||||||||| ||||||| |||||| |||||||||||||||||||  | ||||||||||
Db       1117 CTTGGCTGAAGGTTGAAGTGGTGGATGCTTAAAACTCAATCAAGGTCCTTGAAAAATATG 1176

Qy        721 CTAGTAACCTTGAATGAGAT---------TCAGAGTCGAAATATTGCGAGGCAAGAGCAC 771
              | ||| ||||||||||||||         || |||||||||||||| ||||||| | |||
Db       1177 -TTGTAGCCTTGAATGAGATTCCCATGGATCCCAGTCGAAATATTGTGAGGCAAAATCAC 1235

Qy        772 AATGGAAGAGATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCA 831
              |||| | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1236 AATGAAGGAGATAGCTCCTAGTATAAATATGGATTTATGATATTAACCTATGGTTTGTCA 1295

Qy        832 GCTTTAAATATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCG 891
              | |||| |||||||||||||||||||||||||||||||||||||| | |||| | |||
Db       1296 GTTTTATATATAGCTGTTTGAAACAAAGAATAGAACATATTAGCTA-TTGTTTTCTCGCG 1354

Qy        892 CATGTTATCTTTCTGTTG 909
              ||| ||| |||||||| ||
Db       1355 CATTTTAGCTTTCTGCTG 1372
```

```
RESULT 6
PADAL2
LOCUS       PADAL2                  1753 bp    mRNA    linear   PLN 28-FEB-1995
DEFINITION  P.abies dal2 mRNA.
ACCESSION   X79280
VERSION     X79280.1  GI:695687
KEYWORDS    DAL2 gene.
SOURCE      Picea abies (Norway spruce)
  ORGANISM  Picea abies
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1
  AUTHORS   Tandre,K., Albert,V.A., Sundas,A. and Engstrom,P.
  TITLE     Conifer homologues to genes that control floral development in
            angiosperms
  JOURNAL   Plant Mol. Biol. 27 (1), 69-78 (1995)
  MEDLINE   95170009
   PUBMED   7865797
REFERENCE   2  (bases 1 to 1753)
  AUTHORS   Tandre,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1994) K. Tandre, Uppsala University, Dept of
            Physiological Botany, Villvagen 6, 752 36 Uppsala, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .1753
                     /organism="Picea abies"
                     /mol_type="mRNA"
                     /db_xref="taxon:3329"
                     /clone="3ae"
                     /tissue_type="female cone"
                     /clone_lib="cDNA library spruce female; SF"
                     /dev_stage="before formation of ovule"
     gene            1. .1753
                     /gene="dal2"
     CDS             424. .1092
```

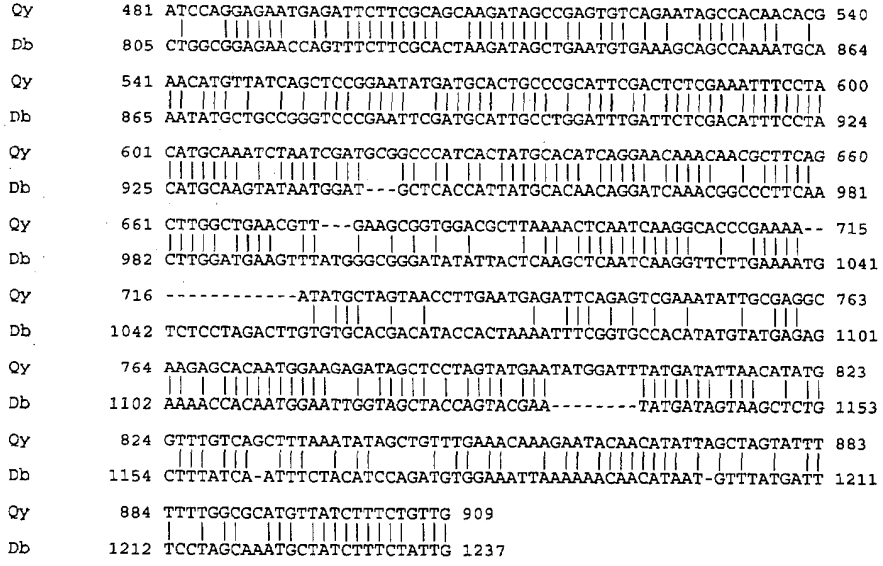
```
                     /gene="dal2"
                     /codon_start=1
                     /product="DAL2 protein"
                     /protein_id="CAA55867.1"
                     /db_xref="GI:695688"
                     /db_xref="SPTREMBL:Q40766"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANHSVKRTIERYKKTCVDNNHGGVISESNSQYWQQEAGKLRQQIE
                     ILQNANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHIL
                     IQENEILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTT
                     LQLG"
     misc_feature    427. .594
                     /gene="dal2"
                     /note="MADS box"
     misc_feature    697. .891
                     /gene="dal2"
                     /note="K box"
BASE COUNT      529 a    340 c    388 g    496 t
ORIGIN

  Query Match            85.2%;  Score 774.8;  DB 8;  Length 1753;
  Best Local Similarity  92.6%;  Pred. No. 2.8e-201;
  Matches  850;  Conservative    0;  Mismatches   57;  Indels   11;  Gaps    3;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db        424 ATGGGCCGTGGGAAGATTGAGATAAAGAGGATTGAGAATACTACGAACCGACAGGTCACT 483

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db        484 TTCTGTAAGCGCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 543

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        544 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAGTTTGCCAACCAC 603

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        604 AGCGTGAAGAGAACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 663

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        664 GTGATTTCGGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 723

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
Db        724 ATTGAAATTTTGCAAAATGCAAATAGGCACTTGATGGGTGACGGGCTTACAGCTTTAAAC 783

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
Db        784 ATCAAGGAACTCAAGCAGCTTGAGGTTCGACTTGAAAAAGGAATCGGCCGAGTACGATCC 843

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
Db        844 AAGAAGAACGAGATGTTGCTTGAGGAGATCGACATCATGCAGAGAAGGGAACACATTCTT 903

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
Db        904 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGCCAGAATAGCCACAACACG 963

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
Db        964 AACATGCTATCAGCTCCCGAATATGATGCATTGCCAGCATTCGATTCTCGAAATTTCCTA 1023

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660
Db       1024 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAGCAAACAACGCTTCAG 1083

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAAATATG 720
Db       1084 CTTGGCTGAAGGTTGAAGTCGTGGATGCTTAAAACTCAATCAAGGTCCTTGAAAAAATATG 1143

Qy        721 CTAGTAACCTTGAATGAGAT---------TCAGAGTCGAAATATTGCGAGGCAAGAGCAC 771
Db       1144 -TTGTAGCCTTGAATGAGATTCCCATGGATCCCAGTCCAAATATTGTGAGGGAAAATCAC 1202

Qy        772 AATGGAAGAGATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCA 831
Db       1203 AATGAAGGAGATAGCTCCTAGTATAAATATGGATTTATGATATTAACCTATGGTTTGTCA 1262

Qy        832 GCTTTAAATATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTGGCG 891
Db       1263 GTTTTATATAAAGCTGTTTGAAACAAAGAATAGAACATATTAGCTA-TTGTTTTCTCGCG 1321

Qy        892 CATGTTATCTTTCTGTTG 909
Db       1322 CATTTTAGCTTTCTGTTG 1339


RESULT 7
PMU69484
LOCUS       PMU69484                1212 bp    mRNA    linear   PLN 18-DEC-1998
DEFINITION  Picea mariana AGAMOUS-like MADS-box transcriptional factor SMADS42D
            mRNA, partial cds.
ACCESSION   U69484
VERSION     U69484.1  GI:4033724
KEYWORDS    .
SOURCE      Picea mariana
  ORGANISM  Picea mariana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 1212)
  AUTHORS   Rutledge,R., Regan,S., Nicolas,O., Fobert,P., Cote,C., Bosnich,W.,
            Kauffeldt,C., Sunohara,G., Seguin,A. and Stewart,D.
  TITLE     Characterization of an AGAMOUS homologue from the conifer black
            spruce (Picea mariana) that produces floral homeotic conversions
            when expressed in Arabidopsis
  JOURNAL   Plant J. 15 (5), 625-634 (1998)
  MEDLINE   98451867
   PUBMED   9778845
REFERENCE   2  (bases 1 to 1212)
  AUTHORS   Rutledge,R.G., Kauffeldt,C., Fobert,P., Cote,C., Bosnich,W. and
            Stewart,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) Canadian Forest Service, Laurentian
            Forestry Centre, 1055 du P.E.P.S., Stainte-Foy, PQ G1V 4C7, Canada
FEATURES             Location/Qualifiers
     source          1. .1212
                     /organism="Picea mariana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3335"
                     /sex="female"
                     /tissue_type="cones"
                     /dev_stage="just prior to pollination"
     CDS             <1. .659
                     /codon_start=3
                     /product="AGAMOUS-like MADS-box transcriptional factor
                     SMADS42D"
                     /protein_id="AAC97159.1"
                     /db_xref="GI:4033725"
                     /translation="KIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVALIVF
                     SSRGRLYEFANHSVKRTIERYKKTCVDNNHGGAISESNSQYWQQEAGKLRQQIEILQN
                     ANRHLMGDGLTALNIKELKQLEVRLEKGIGRVRSKKNEMLLEEIDIMQRREHILIQEN
                     EILRSKIAECQNSHNTNMLSAPEYDALPAFDSRNFLHANLIDAAHHYAHQEQTTLQLG
                     "
BASE COUNT      391 a    216 c    268 g    337 t
ORIGIN

  Query Match            85.0%;  Score 772.8;  DB 8;  Length 1212;
  Best Local Similarity  93.1%;  Pred. No. 9.7e-201;
  Matches  845;  Conservative    0;  Mismatches   52;  Indels   11;  Gaps    3;

Qy         11 GGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTTTCTGCAAGC 70
Db          1 GGAAGATTGAGATAAAGAGGATTGAGAATACTACGAACCGACAGGTCACTTTCTGTAAGC 60
```

```
Qy         71 GCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATGCAGAAGTGG 130
Db         61 GCCGAAATGGTTTATTGAAGAAGGCGTATGAATTATCAGTTCTTTGTGATGCAGAAGTGG 120

Qy        131 CCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACAGCGTGAAGA 190
Db        121 CCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAGTTTGCCAACCACAGCGTGAAGA 180

Qy        191 GGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGGCGATATCAG 250
Db        181 GAACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGGCGATTTCAG 240

Qy        251 AGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGATTGACATTT 310
Db        241 AGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGATTGAAATTT 300

Qy        311 TGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACATTAAGGAAC 370
Db        301 TGCAAAATGCAAATAGGCACTTGATGGGTGATGGGCTTACAGCTTTAAACATCAAGGAAC 360

Qy        371 TCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCAAAAAGAACG 430
Db        361 TCAAGCAGCTTGAGGTTCGACTTGAAAAAGGAATCGGCCGAGTACGATCCAAGAAGAACG 420

Qy        431 AGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTATCCAGGAGA 490
Db        421 AGATGTTGCTTGAGGAGATCGACATCATGCAGAGAAGGGAACACATTCTAATCCAGGAGA 480

Qy        491 ATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACGAACATGTTAT 550
Db        481 ATGAGATTCTTCGCAGCAAGATAGCCGAGTGCCAGAATAGCCACAACACGAACATGCTAT 540

Qy        551 CAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTACATGCAAATC 610
Db        541 CAGCTCCCGAATATGATGCATTGCCAGCATTCGATTCTCGAAATTTCCTACATGCAAATC 600

Qy        611 TAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAGCTTGGCTGAA 670
Db        601 TAATCGATGCGGCCCATCACTATGCACATCAGGAGCAAACAACGCTTCAGCTTGGCTGAA 660

Qy        671 CGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATGCTAGTAACCT 730
Db        661 GGTTGAAGTGGTGGATGCTTAAAACTCAATCAAGGTCCTTGAAAAATATG-TTGTAGCCT 719

Qy        731 TGAATGAGAT---------TCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGAG 781
Db        720 TGAATGAGATTCCCATGGATCCCAGTCGAAATATTGTGAGGCAAAATCACAATGAAGGAG 779

Qy        782 ATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAATA 841
Db        780 ATAGCTCCTAGTATAAATATGGATTTATGATATTAACCTATGGTTTGTCAGTTTTATATA 839

Qy        842 TAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATCT 901
Db        840 TAGCTGTTTGAAACAAAGAATAGAACATATTAGCTA-TTGTTTTCTCGCGCATTTTAGCT 898

Qy        902 TTCTGTTG 909
Db        899 TTCTGTTG 906
```

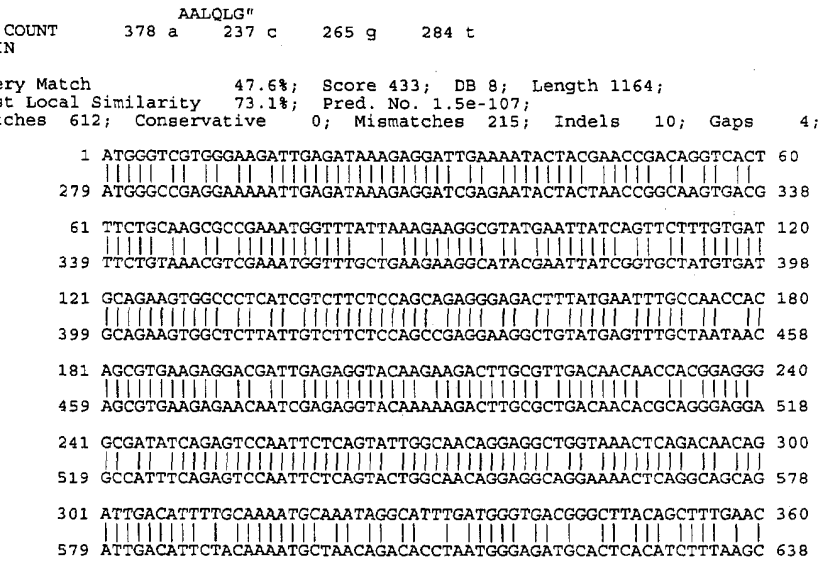
```
RESULT 8
AY114304
LOCUS       AY114304               1372 bp    mRNA    linear   PLN 29-APR-2003
DEFINITION  Ginkgo biloba AGAMOUS-like MADS-box transcription factor (GBM5)
            mRNA, complete cds.
ACCESSION   AY114304
VERSION     AY114304.1  GI:30230269
KEYWORDS    .
SOURCE      Ginkgo biloba (maidenhair tree)
  ORGANISM  Ginkgo biloba
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Ginkgophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
REFERENCE   1  (bases 1 to 1372)
  AUTHORS   Jager,M., Hassanin,A., Manuel,M., Le Guyader,H. and Deutsch,J.
  TITLE     MADS-Box Genes in Ginkgo biloba and the Evolution of the AGAMOUS
            Family
  JOURNAL   Mol. Biol. Evol. 20 (5), 842-854 (2003)
   PUBMED   12679535
REFERENCE   2  (bases 1 to 1372)
  AUTHORS   Jager,M., Hassanin,A., Manuel,M., Le Guyader,H. and Deutsch,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2002) Universite Paris VI, 9 quai Saint Bernard,
            Paris 75252 cedex05, France
FEATURES             Location/Qualifiers
     source          1. .1372
                     /organism="Ginkgo biloba"
                     /mol_type="mRNA"
                     /db_xref="taxon:3311"
                     /tissue_type="stamen"
     gene            1. .1372
                     /gene="GBM5"
     5'UTR           1. .324
                     /gene="GBM5"
     CDS             325. .990
                     /gene="GBM5"
                     /codon_start=1
                     /product="AGAMOUS-like MADS-box transcription factor"
                     /protein_id="AAM76208.1"
                     /db_xref="GI:30230270"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANNSVKRTIDRYKKTCADNSQGGAISECNSQYWQQEAGKLRQQID
                     ILQNANRHLMGDALTSLSVKELKQLEIRLERGISRVRSKKNEMLLEEIEIMQRREHIL
                     LAENQFLRTKIAECESSQNANMLPGPEFDALPGFDSRHFLHASIMDAHHYAQQDQTAL
                     QLG"
     3'UTR           991. .1372
                     /gene="GBM5"
BASE COUNT      439 a    270 c    275 g    388 t
ORIGIN
```

Query Match 50.1%; Score 455.6; DB 8; Length 1372;
Best Local Similarity 73.1%; Pred. No. 9.4e-114;
Matches 677; Conservative 0; Mismatches 219; Indels 30; Gaps 6;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db        325 ATGGGCCGTGGGAAGATTGAGATAAAGAGGATTGAGAACACTACAAACCGACAGGTCACT 384

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db        385 TTCTGCAAGCGTCGAAATGGTCTGTTGAAGAAAGCGTATGAATTATCCGTGCTTTGTGAT 444

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        445 GCGGAAGTGGCTCTCATCGTCTTCTCCAGTCGCGGGAGACTCTATGAGTTTGCTAATAAC 504

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        505 AGTGTGAAGAGAACAATCGACAGATACAAGAAGACTTGCGCTGACAACTCTCAGGGTGGA 564

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        565 GCCATTTCAGAGTGTAATTCTCAGTACTGGCAACAGGAGGCAGGAAAACTGAGACAGCAA 624

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
Db        625 ATTGATATTCTGCAAAATGCAAATAGACACTTGATGGGGGACGCGCTTACATCTTTAAGT 684

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
Db        685 GTAAAGGAGCTTAAGCAGCTAGAAATTCGACTTGAGAGGGGCATTAGCAGGGTTCGATCA 744

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
Db        745 AAGAAGAATGAAATGTTGCTTGAGGAGATAGAGATTATGCAAAGAAGGGAACACATATTA 804
```



RESULT 9
AF492455
LOCUS AF492455 1164 bp mRNA linear PLN 02-JUL-2002
DEFINITION Cycas edentata MADS-box transcription factor (AG) mRNA, complete cds.
ACCESSION AF492455
VERSION AF492455.1 GI:21667495
KEYWORDS .
SOURCE Cycas edentata
ORGANISM Cycas edentata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Cycadophyta; Cycadales; Cycadaceae; Cycas.
REFERENCE 1 (bases 1 to 1164)
AUTHORS Zhang,P., Tan,T.W.H., Pwee,K.H., Lim,S.H. and Kumar,P.P.
TITLE CyAG, the orthologous gene of AGAMOUS, from Cycas edentata
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1164)
AUTHORS Zhang,P., Tan,T.W.H., Pwee,K.H., Lim,S.H. and Kumar,P.P.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2002) Department of Biological Sciences, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore
FEATURES Location/Qualifiers
source 1. .1164
/organism="Cycas edentata"
/mol_type="mRNA"
/db_xref="taxon:179041"
gene 1. .1164
/gene="AG"
CDS 279. .953
/gene="AG"
/note="CyAG protein"
/codon_start=1
/product="MADS-box transcription factor"
/protein_id="AAM74074.1"
/db_xref="GI:21667496"
/translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
LIVFSSRGRLYEFANNSVKRTIERYKKTCADNTQGGAISESNSQYWQQEAGKLRQQID
ILQNANRHLMGDALTSLSVKELKQLEIRLERGLSRVRSKKNEMLLEEIEIMQRREHIL
LAENQFLRTKIAEYESNQNTNVLIPGPEFDALPAFDSRNFLHANLIEAAAHHYTQQDQ



RESULT 10
GGN132209
LOCUS GGN132209 1154 bp mRNA linear PLN 23-NOV-1999
DEFINITION Gnetum gnemon mRNA for putative MADS domain transcription factor GGM3.
ACCESSION AJ132209
VERSION AJ132209.1 GI:5019430
KEYWORDS ggm3 gene; MADS domain transcription factor.

```
SOURCE      Gnetum gnemon
  ORGANISM  Gnetum gnemon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Gnetophyta; Gnetopsida; Gnetales; Gnetaceae; Gnetum.
REFERENCE   1
  AUTHORS   Winter,K.U., Becker,A., Munster,T., Kim,J.T., Saedler,H. and
            Theissen,G.
  TITLE     MADS-box genes reveal that gnetophytes are more closely related to
            conifers than to flowering plants
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (13), 7342-7347 (1999)
  MEDLINE   99307411
   PUBMED   10377416
REFERENCE   2  (bases 1 to 1154)
  AUTHORS   Theissen,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Theissen G., Molecular Plant Genetics,
            Max-Planck-Institut for Breeding Reserch, Carl-von-Linne Weg 10,
            50829 Koeln, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1154
                     /organism="Gnetum gnemon"
                     /mol_type="mRNA"
                     /sub_species="gnemon"
                     /db_xref="taxon:3382"
     gene            1. .1154
                     /gene="ggm3"
     CDS             88. .831
                     /gene="ggm3"
                     /codon_start=1
                     /product="putative MADS domain transcription factor GGM3"
                     /protein_id="CAB44449.1"
                     /db_xref="GI:5019431"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEFANNSVKRTIERYRKTCADNNQGGAIAESNAQYWQQEAVKLKQQID
                     VLNNQIRHYMGECLQSMTIKELKQLEGKLEKGLGRVRSKRNEKLLEDIDTLQRREDNL
                     IRENEYIRNKIAECQSHQHANMLTAAAVEYDAIPAAYDSRNFMHANLIEAAAAHHHYA
                     QQEQTALHLGSEHKYSVHYPDPQAHKFQT"
BASE COUNT      413 a    238 c    255 g    248 t
ORIGIN

  Query Match           36.1%;  Score 328.6;  DB 8;  Length 1154;
  Best Local Similarity   69.9%;  Pred. No. 6.2e-79;
  Matches  480;  Conservative    0;  Mismatches  189;  Indels   18;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||||| || ||| |||||||||||||||||||| || || || || ||| ||||||| |||
Db         88 ATGGGACGCGGCAAGATCGAGATAAAGAGGATCGAGAACACAACCAACAGACAGGTAACT 147

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||||||| | ||||| |  | |||||||| || || ||||| ||||| || |||
Db        148 TTCTGCAAGCGCAGGAATGGCCTGCTGAAGAAGGCCTACGAGTTATCGGTTCTCTGCGAT 207

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || |||||||||||| |||||||| |||| | || |||||||| |||||| ||
Db        208 GCTGAGGTCGCCCTCATCGTTTTCTCCAGCCGAGGCCGTCTCTATGAATTCGCCAACAAC 267

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || || ||||| ||||| |||||||| |  |||||||| | |||||||| || ||||||
Db        268 AGTGTCAAGAGAACGATCGAGAGGTATAGAAAGACTTGTGCGGACAACAATCAGGGAGGA 327

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              || ||  |||| |||||  | |||||||||||| ||||| |||| |||||  | || ||
Db        328 GCAATTGCAGAATCCAACGCACAGTATTGGCAGCAGGAAGCTGTGAAACTGAAGCAGCAA 387

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              || ||| |  |  | ||  |  | || |  ||||| ||  | |||    |  |||
Db        388 ATAGACGTCCTCAACAACCAGATCCGACACTACATGGGAGAATGCCTTCAGTCCATGACA 447

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              || || ||||| ||||| || |||| |  ||| || |||||| |  |  |||| ||||||
Db        448 ATCAAAGAACTGAAGCAGCTGGAGGGTAAACTCGAGAAAGGACTTGGTAGAGTTCGATCC 507

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
```

```
              |||| ||| ||||   |||| || || || |||| | |||||||||||||| | |  ||
Db        508 AAAAGGAATGAGAAACTGCTGGAGGACATAGACACCTTGCAGAGAAGGGAAGATAACCTG 567

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              || |  |||||||||     |||| | |||||||| || || ||||    ||  | |
Db        568 ATTCGAGAGAATGAGTACATTCGGAACAAGATAGCAGAATGCCAGAGCCATCAGCATGCA 627

Qy        541 AACATGTTATCAGCTCC---------GGAATATGATGCACTGCCCGCATTCGACTCTCGA 591
              |||||||  | ||  |         | |  |||    ||   | ||||  || ||  |
Db        628 AACATGTTGACTGCAGCTGCTGTAGAGTATGATGCCATACCTGCAGCATATGATTCCAGG 687

Qy        592 AATTTCCTACATGCAAATCTAATCGATGCGGC---------CCATCACTATGCACATCAG 642
              || ||| | |||||||| || || || || ||          |||||  |||||  || ||
Db        688 AACTTCATGCATGCAAACCTGATAGAAGCAGCAGCTGCTCATCATCATTATGCTCAACAA 747

Qy        643 GAACAAACAACGCTTCAGCTTGGCTGA 669
              || |||||  |||||||  |||||  | |
Db        748 GAGCAAACTGCGCTTCATCTTGGATCA 774
```

```
RESULT 11
AB050644
LOCUS       AB050644                 912 bp    mRNA    linear   PLN 31-OCT-2001
DEFINITION  Magnolia praecocissima mRNA for putative MADS-domain transcription
            factor MpMADS2, partial cds.
ACCESSION   AB050644
VERSION     AB050644.1  GI:16549059
KEYWORDS    .
SOURCE      Magnolia praecocissima
  ORGANISM  Magnolia praecocissima
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; magnoliids; Magnoliales;
            Magnoliaceae; Magnolia.
REFERENCE   1
  AUTHORS   Ito,M., Shiobara,S., Tanabe,Y. and Hasebe,M.
  TITLE     Organ identities in Magnolian flower
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 912)
  AUTHORS   Ito,M., Shiobara,S. and Hasebe,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Motomi Ito, University of Tokyo, Department
            of Systems Sciences, Graduate School of Arts and Sciences; 3-8-1
            Komaba, Meguro-ku, Tokyo 153-8902, Japan
            (E-mail:cmito@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5454-6638,
            Fax:81-3-5454-6638)
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="Magnolia praecocissima"
                     /mol_type="mRNA"
                     /db_xref="taxon:81865"
     gene            1. .912
                     /gene="MpMADS2"
     CDS             <1. .627
                     /gene="MpMADS2"
                     /codon_start=1
                     /product="putative MADS-domain transcription factor
                     MpMADS2"
                     /protein_id="BAB70737.1"
                     /db_xref="GI:16549060"
                     /translation="NRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSSRGRLYEYAN
                     NSVRNTIDRYKKACADSSSLGCVSEANSQYYQQESSKLRQQIALLQNANRHLMGEALS
                     SMTVKELKQLENRLEKGISRIRSKKNELLFAEIEYMQKREVDLQNDNMYLRAKITENE
                     RAQQQMGMLPAPEYDVMPGFDSRNFLQVNLMDSSHHYSHQEQTALQLG"
     polyA_signal    902. .911
                     /gene="MpMADS2"
BASE COUNT      282 a    184 c    199 g    247 t
ORIGIN

  Query Match           33.4%;  Score 303.8;  DB 8;  Length 912;
  Best Local Similarity   67.7%;  Pred. No. 3.9e-72;
  Matches  441;  Conservative    0;  Mismatches  207;  Indels    3;  Gaps    1;

Qy         46 AACCGACAGGTCACTTTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTA 105
```

```
Db          1 AACAGGCAAGTCACTTTTTGCAAGCGCCGTAATGGCCTCCTCAAGAAGGCCTATGAACTG 60

Qy        106 TCAGTTCTTTGTGATGCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTAT 165

Db         61 TCTGTGCTTTGTGATGCTGAGGTGGCTCTCATCGTCTTCTCCAGCCGCGGCCGCCTCTAC 120

Qy        166 GAATTTGCCAACCACAGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGAC 225

Db        121 GAATACGCCAATAACAGTGTTAGGAATACAATCGATAGGTATAAGAAGGCATGTGCAGAT 180

Qy        226 AACAACCACGGAGGGGCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGT 285

Db        181 TCCTCAAGTTTAGGATGCGTATCGGAAGCCAATTCTCAGTATTACCAGCAAGAATCATCA 240

Qy        286 AAACTCAGACAACAGATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGG 345

Db        241 AAACTGCGCCAACAAATAGCGCTTTTACAGAATGCAAACAGGCATTTGATGGGTGAGGCC 300

Qy        346 CTTACAGCTTTGAACATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATC 405

Db        301 CTCAGTTCCATGACTGTTAAGGAGCTCAAGCAGCTGGAGAATAGATTGGAGAAAGGCATT 360

Qy        406 AGCCGAGTGCGATCCAAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGA 465

Db        361 AGCAGAATCAGATCTAAGAAGAACGAGTTGCTGTTTGCTGAAATCGAGTACATGCAGAAA 420

Qy        466 AGGGAACACATACTTATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAG 525

Db        421 AGGGAGGTTGATTTGCAAAATGATAATATGTACCTCCGTGCTAAGATAACTGAGAACGAA 480

Qy        526 AATAGC---CACAACACGAACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTC 582

Db        481 AGAGCCCAACAGCAAATGGGCATGTTGCCTGCTCCTGAATACGATGTAATGCCTGGATTT 540

Qy        583 GACTCTCGAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAG 642

Db        541 GATTCTCGGAACTTCCTACAAGTAAATCTAATGGATTCCAGTCACCATTACTCTCACCAG 600

Qy        643 GAACAAACAACGCTTCAGCTTGGCTGAACGTTGAAGCGGTGGACGCTTAAA 693

Db        601 GAGCAAACAGCCCTCCAACTTGGCTAAAAGGTAATGTCGTGGATGGTCCAA 651
```

RESULT 12
AF373604
LOCUS AF373604 942 bp mRNA linear PLN 21-JUN-2002
DEFINITION Vitis vinifera MADS-box protein 5 (MADS5) mRNA, complete cds.
ACCESSION AF373604
VERSION AF373604.1 GI:20385589
KEYWORDS .
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 942)
AUTHORS Boss,P.K., Sensi,E., Hua,C., Davies,C. and Thomas,M.R.
TITLE Cloning and characterization of grapevine (Vitis vinifera L.) MADS-box genes expressed during inflorescence and berry development
JOURNAL Plant Sci. 162 (6), 887-895 (2002)
REFERENCE 2 (bases 1 to 942)
AUTHORS Boss,P.K., Sensi,E., Hua,C., Davies,C. and Thomas,M.R.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) Plant Industry, CSIRO, Hartley Grove, Urrbrae, SA 5064, Australia
FEATURES Location/Qualifiers
source 1. .942
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
gene 1. .942
/gene="MADS5"
CDS 71. .742
/gene="MADS5"
/note="VvMADS5"
/codon_start=1
/product="MADS-box protein 5"
/protein_id="AAM21345.1"
/db_xref="GI:20385590"
/translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
LIVFSSRGRVYEYSNNNIKSTIDRYKKASSDSTNGGFTMEINAQYYQQESAKLRQQIQ
MLQNSNRHLMGDSLASLTVKELKQLENRLERGITRIRSKKHELLLAEIEYLQKREIEL
ENESVYLRTKIAEVERLQQANMVSTHEFNAIQALVSRNFFQPNMIEGGSTGYPLHDKK
VLHLG"
BASE COUNT 300 a 183 c 216 g 243 t
ORIGIN

Query Match 30.8%; Score 279.6; DB 8; Length 942;
Best Local Similarity 64.3%; Pred. No. 1.7e-65;
Matches 436; Conservative 0; Mismatches 239; Indels 3; Gaps 1;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db         71 ATGGGGAGAGGAAAGATCGAGATCAAGAGGATCGAAAACACGACCAACCGTCAGGTCACA 130

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        131 TTCTGCAAGCGAAGGAATGGGCTTTTGAAGAAGGCTTATGAATTATCAGTGCTATGTGAT 190

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        191 GCAGAAGTTGCCCTCATCGTCTTCTCCAGCCGCGGTCGAGTCTATGAGTACTCAAACAAC 250

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        251 AACATAAAATCAACCATAGATAGGTACAAGAAGGCCAGCTCAGATAGTACAAATGGAGGC 310

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        311 TTTACCATGGAGATCAATGCCCAATATTACCAGCAAGAATCAGCAAAGCTGCGCCAGCAA 370

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        371 ATACAGATGCTGCAGAATTCTAACAGGCACTTAATGGGTGATTCCTTGGCTTCCTTGACT 430

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        431 GTGAAGGAGCTAAAGCAGCTCGAGAACAGGCTTGAACGAGGCATCACAAGAATCAGGTCG 490

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        491 AAGAAGCATGAGTTGCTGTTGGCTGAGATTGAATACTTGCAGAAAAGGGAAATTGAGCTG 550

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        551 GAAAATGAAAGCGTATATCTCCGAACCAAGATTGCAGAAGTGGAGAGGCTTCAGCAAGCA 610

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        611 AACATGGTATCAACACATGAGTTCAATGCCATCCAGGCATTAGTTTCTCGCAATTTCTTT 670

Qy        601 CATGCAAATCTAATCGA---TGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTT 657

Db        671 CAGCCCAATATGATTGAGGGTGGATCCACAGGCTACCCACTTCATGATAAGAAGGTCCTC 730

Qy        658 CAGCTTGGCTGAACGTTG 675

Db        731 CATCTCGGGTAAATGATG 748
```

RESULT 13
AF234617
LOCUS AF234617 1009 bp mRNA linear PLN 12-FEB-2002
DEFINITION Phalaenopsis equestris MADS-box transcription factor (MADS1) mRNA,

```
            complete cds.
ACCESSION   AF234617
VERSION     AF234617.1  GI:18650788
KEYWORDS    .
SOURCE      Phalaenopsis equestris
  ORGANISM  Phalaenopsis equestris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
            Epidendroideae; higher Epidendroideae; Vandeae; Aeridinae;
            Phalaenopsis.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Chen,H.H., Liu,C.C., Tsuei,S.W. and Chen,W.H.
  TITLE     Phalaenopsis equestris MADS-box protein (MADS1)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1009)
  AUTHORS   Chen,H.H. and Tsuei,S.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Institute of Biotechnology, National Cheng
            Kung University, Tainan, Taiwan 701
REFERENCE   3  (bases 1 to 1009)
  AUTHORS   Chen,W.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Department of Horticulture, Taiwan Sugar
            Research Institute, Tainan 702, Taiwan
REFERENCE   4  (bases 1 to 1009)
  AUTHORS   Liu,C.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Department of Biology, National Cheng Kung
            University, Tainan 701, Taiwan
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Phalaenopsis equestris"
                     /mol_type="mRNA"
                     /db_xref="taxon:78828"
                     /clone="PeMADS1"
                     /tissue_type="flower bud"
     gene            1. .1009
                     /gene="MADS1"
                     /note="homeotic gene"
     CDS             112. .831
                     /gene="MADS1"
                     /codon_start=1
                     /product="MADS-box transcription factor"
                     /protein_id="AAL76415.1"
                     /db_xref="GI:18650789"
                     /translation="MDSSSMEPKEKMGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAY
                     ELSVLCDAEVALIIFSTRGRLYEYANNSVKGTIERYKKASTDNSNTGSISEANSQYYQ
                     QEATKLRQQITNLQNSNRNLLGDALTTMSLRDLKQLETRLEKGINKIRAKKNELLHAE
                     IDYMQKREMELQTDNMFLRNKISDNERAQQQHQHMSILPSTSTEYEVMPPFDSRSFLH
                     VNLMDPNDRYSHQQQTALQLG"
BASE COUNT      337 a    188 c    235 g    249 t
ORIGIN
```

```
  Query Match             30.7%;  Score 279;  DB 8;  Length 1009;
  Best Local Similarity   65.4%;  Pred. No. 2.5e-65;
  Matches  449;  Conservative    0;  Mismatches  220;  Indels   18;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | ||||||||| ||||| ||||| || || || || || ||| |||| || ||
Db        145 ATGGGGAGGGGGAAGATAGAGATCAAGAGAATAGAGAACACAACAAACAGACAAGTTACC 204

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              || |||||||||||| ||||||  |  | |||||||| |||||  | || ||||| ||||||
Db        205 TTTTGCAAGCGCCGTAATGGCCTCCTGAAGAAGGCTTATGAGCTCTCTGTTCTGTGTGAT 264

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              v  || || || ||||| ||| |||||||| | | | ||  | || |||||||| ||| ||| ||
Db        265 GCTGAGGTTGCCCTAATCATCTTCTCTACCCGTGGCCGCCTCTATGAATATGCAAACAAC 324

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||||||||  | || |||||  | ||||| ||| |  ||  ||| ||    | |    ||
Db        325 AGCGTGAAAGGAACCATTGAACGATACAAAAAGGCAAGCACTGATAATTCCAATACTGGA 384

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||||| ||  | ||||| |||||||  ||||| |||||    |||||  | || ||
Db        385 TCAATATCTGAAGCAAATTCGCAGTATTACCAACAAGAGGCCACAAAACTGCGTCAGCAA 444

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||    |  || || ||| | || ||| ||||| |||||||  |  || ||  | ||| |
Db        445 ATTACAAACTTACAGAATTCCAACAGGAATTTGCTGGGTGATGCTCTCACCACCATGAGC 504

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              | | ||| || |||||||| ||       || | || ||||| |||| |  | |  || |
Db        505 CTGAGGGACCTGAAGCAACTGGAAACAAGATTGGAGAAAGGCATCAACAAAATAAGAGCT 564

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||||| ||  || ||| ||  ||||| |||  |||||||| |||||||      |||
Db        565 AAGAAGAATGAACTGCTGCATGCTGAGATTGACTACATGCAGAAAAGGGAAATGGAACTC 624

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA----GTGTCAGAATAGCCACAA 536
                 || ||   | |||| ||||  ||||||  | ||     |    || | |||  ||
Db        625 CAAACTGACAACATGTTTCTGCGCAATAAGATATCTGATAATGAAAGAGCACAGCAGCAG 684

Qy        537 CACGAACATGTTATCAGCT-------------CCGGAATATGATGCACTGCCCGCATTC 582
              ||  | ||| | | ||  |                | || ||||| |   ||||  ||||
Db        685 CATCAGCATATGAGCATATTGCCATCAACAAGCACAGAGTATGAAGTGATGCCTCCATTT 744

Qy        583 GACTCTCGAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAG 642
              || ||  |||  || || |||| |||||||| ||| |     | |  ||| |||| |||
Db        745 GATTCCAGAAGCTTTCTTCATGTCAATCTAATGGATCCCAATGACCGTTATTCCCACCAG 804

Qy        643 GAACAAACAACGCTTCAGCTTGGCTGA 669
              | |||||| | || || ||||| |||
Db        805 CAGCAAACAGCTCTGCAACTTGGGTGA 831
```

```
RESULT 14
AF022377
LOCUS       AF022377               1142 bp    mRNA    linear   PLN 17-AUG-2001
DEFINITION  Cucumis sativus agamous-like putative transcription factor (CAG1)
            mRNA, complete cds.
ACCESSION   AF022377
VERSION     AF022377.1  GI:4103341
KEYWORDS    .
SOURCE      Cucumis sativus (cucumber)
  ORGANISM  Cucumis sativus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE   1  (bases 1 to 1142)
  AUTHORS   Perl-Treves,R., Kahana,A., Rosenman,N., Xiang,Y. and Silberstein,L.
  TITLE     Expression of multiple AGAMOUS-like genes in male and female
            flowers of cucumber (Cucumis sativus L.)
  JOURNAL   Plant Cell Physiol. 39 (7), 701-710 (1998)
  MEDLINE   98399593
   PUBMED   9729894
REFERENCE   2  (bases 1 to 1142)
  AUTHORS   Perl-Treves,R., Rosenman,N., Kahana,A. and Silberstein,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-1997) Life Science, Bar-Ilan University,
            Ramat-Gan 52900, Israel
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Cucumis sativus"
                     /mol_type="mRNA"
                     /db_xref="taxon:3659"
                     /tissue_type="flower"
                     /dev_stage="bud"
     gene            1. .1142
                     /gene="CAG1"
     CDS             173. .850
                     /gene="CAG1"
                     /codon_start=1
                     /product="agamous-like putative transcription factor"
                     /protein_id="AAD01742.1"
                     /db_xref="GI:4103342"
```

```
                /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                LIVFSSRGRLYEYSNNSIKTTIERYKKACSDSSATSSVTELNTQYYQQESAKLRQQIQ
                MLQNSNRHLMGDSLSALTVKELKQLENRLERGITRIRSKKHEMLLAEIEYLQKREIEL
                ENENVCIRTKIAEVERVQQANMVSGQELNAIQALANSRNFFSPNIMEPAGPVSYSHQD
                KKMLHLG"
BASE COUNT      413 a    193 c    209 g    327 t
ORIGIN

  Query Match            30.5%;  Score 276.8;  DB 8;  Length 1142;
  Best Local Similarity  65.2%;  Pred. No. 1e-64;
  Matches  442;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db        173 ATGGGGAGAGGAAAGATAGAGATAAAGAGAATAGAGAACACAACAAATAGACAAGTTACA 232

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db        233 TTCTGTAAGAGAAGAAATGGACTTTTGAAAAAAGCTTATGAACTCTCTGTTCTTTGTGAT 292

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        293 GCTGAAGTTGCTCTCATTGTCTTCTCCAGCCGTGGCCGTCTCTATGAATACTCCAATAAC 352

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        353 AGCATCAAAACAACTATTGAGAGGTACAAGAAGGCTTGTTCTGATAGCTCAGCTACTAGC 412

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        413 TCTGTCACTGAACTAAATACTCAATATTATCAGCAAGAATCGGCTAAACTGCGTCAACAG 472

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
Db        473 ATACAAATGCTTCAGAATTCCAACAGGCACTTGATGGGGACTCCTTGAGTGCTCTTACT 532

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
Db        533 GTCAAAGAACTCAAGCAGCTTGAAAATAGGCTTGAAAGAGGCATCACTAGAATCAGATCA 592

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
Db        593 AAGAAGCACGAAATGTTGCTAGCAGAAATTGAGTACCTTCAGAAAAGGGAGATTGAGCTG 652

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
Db        653 GAGAACGAAAATGTGTGTATTAGAACCAAGATAGCAGAAGTAGAGAGGGTTCAACAAGCA 712

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCG---ACTCTCGAAATTTC 597
Db        713 AACATGGTATCTGGACAAGAACTGAATGCAATACAAGCATTGGCTAACTCTCGCAATTTC 772

Qy        598 CTACATGCAAATCTAATCGA------TGCGGCCCATCACTATGCACATCAGGAACAAACA 651
Db        773 TTCTCTCCCAATATCATGGAACCTGCTGGACCTGTTTCTTACTCTCATCAAGACAAGAAA 832

Qy        652 ACGCTTCAGCTTGGCTGA 669
Db        833 ATGCTTCATCTTGGGTGA 850

RESULT 15
AY178837
LOCUS       AY178837               1001 bp    mRNA    linear   PLN 15-JAN-2003
DEFINITION  Momordica charantia mads-box transcription factor mRNA, complete
            cds.
ACCESSION   AY178837
VERSION     AY178837.1  GI:27763669
KEYWORDS    .
SOURCE      Momordica charantia (balsam pear)
  ORGANISM  Momordica charantia
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Momordica.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Yang,M., Lan,L., Zeng,Y., Zhang,M., Bai,J., Miao,C., Cai,Y.,
            Lai,J., Xu,Y. and Chen,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2002) Sichuan University, #29 Wangjiang Road,
            Chengdu, Sichuan 610064, P. R. China
FEATURES             Location/Qualifiers
     source          1..1001
                     /organism="Momordica charantia"
                     /mol_type="mRNA"
                     /db_xref="taxon:3673"
     CDS             51..734
                     /codon_start=1
                     /product="mads-box transcription factor"
                     /protein_id="AAO20104.1"
                     /db_xref="GI:27763670"
                     /translation="MGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVA
                     LIVFSSRGRLYEYSNNSIKTTIGRYKKACSDSSATSSVTELNTQYYQQESAKLRQQIQ
                     MLQNSNSNLVRHLMGDSLSALTVKELKQLENRLERGITRIRSKKHEMLLAEIEYLQKR
                     EIELENENVCIRTKIAEVERLQQANMVSGQELNAIQALASRNFFTPNMMEGGAVTFSH
                     QDKKMLHLG"
BASE COUNT      337 a    162 c    239 g    263 t
ORIGIN

  Query Match            29.3%;  Score 266;  DB 8;  Length 1001;
  Best Local Similarity  64.2%;  Pred. No. 9e-62;
  Matches  439;  Conservative    0;  Mismatches  230;  Indels   15;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db         51 ATGGGGAGAGGGAAGATCGAGATTAAGCGGATCGAGAACACGACAAATCGGCAGGTGACG 110

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db        111 TTCTGCAAGAGGAGGAATGGACTGCTGAAGAAAGCTTATGAACTGTCTGTTCTTTGTGAT 170

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        171 GCTGAAGTTGCCCTCATCGTCTTCTCCAGCCGTGGTCGCCTCTATGAATACTCCAATAAC 230

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        231 AGCATCAAAACTACCATTGGAGAGGTACAAGAAGGCTTGTTCTGATAGCTCAGCTACTAGC 290

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        291 TCTGTCACTGAACTCAATACTCAATATTATCAGCAAGAATCTGCAAAGCTGCGCCAACAA 350

Qy        301 ATTGACATTTTGCAAAATGCAAA------------TAGGCATTTGATGGGTGACGGGCTT 348
Db        351 ATACAAATGCTACAGAATTCCAACAGCAATCTTGTTAGGCACTTGATGGGGACTCCTTG 410

Qy        349 ACAGCTTTGAACATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGC 408
Db        411 AGTGCTCTTACAGTGAAGGAATTGAAGCAGCTTGAGAATAGGCTTGAAAGAGGCATCACT 470

Qy        409 CGAGTGCGATCCAAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGG 468
Db        471 AGGATCAGATCAAAGAAGCATGAAATGTTGCTAGCAGAAATTGAGTACCTGCAGAAAAGA 530

Qy        469 GAACACATACTTATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAAT 528
Db        531 GAGATCGAGCTGGAGAATGAAAATGTGTGTATTAGAACCAAGATAGCAGAAGTGGAGAGG 590

Qy        529 AGCCACAACACGAACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCT 588
Db        591 CTTCAGCAAGCCAACATGGTTTCTGGGCAAGAACTGAATGCAATTCAGGCATTGGCTTCT 650

Qy        589 CGAAATTTCCTACATGCAAATCTAATCGA---TGCGGCCCATCACTATGCACATCAGGAA 645
Db        651 CGAAATTTCTTCACTCCTAATATGATGGAAGGTGGAGCTGTCACTTTCTCACACCAAGAC 710
```

```
Qy        646 CAAACAACGCTTCAGCTTGGCTGA 669
              | |   | |||||| ||||| |||
Db        711 AAGAAGATGCTTCATCTTGGGTGA 734
```

Search completed: January 21, 2004, 22:14:26
Job time : 3322.05 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 234.397 Seconds
(without alignments)
10468.497 Million cell updates/sec

Title: US-09-936-869-3
Perfect score: 909
Sequence: 1 atgggtcgtgggaagattga.....cgcatgttatctttctgttg 909

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	909	100.0	909	21	AAA99364	Plant PrAG1 promot
2	260	28.6	1201	21	AAZ57947	Poplar floral home
3	260	28.6	1219	22	AAF85398	Nucleotide sequenc
4	260	28.6	1219	24	ABK88489	Poplar floral home
5	259.8	28.6	723	22	AAF85399	Nucleotide sequenc
6	255.6	28.1	777	21	AAC43464	Arabidopsis thalia
7	255.6	28.1	959	20	AAV99860	Arabidopsis AGL5 c
8	255.6	28.1	959	24	ABQ79456	Arabidopsis AGL5 e
9	255.6	28.1	959	25	ABX11511	DNA encoding Arabi
10	251.4	27.7	1157	21	AAZ50201	Petunia hybrida ne
11	251.4	27.7	1157	21	AAZ35494	Petunia nectary-sp
12	248.2	27.3	714	22	AAF85402	Nucleotide sequenc
13	248.2	27.3	1115	21	AAZ57949	Poplar floral home
14	248.2	27.3	1159	22	AAF85401	Nucleotide sequenc
15	248.2	27.3	1159	24	ABK88491	Poplar floral home
16	248.2	27.3	1297	24	ABS55501	Sweetgum cDNA enco
17	245.2	27.0	706	21	AAC55965	Eucalyptus grandis
18	245	27.0	896	20	AAV99859	Arabidopsis AGL1 c
19	245	27.0	896	24	ABQ79455	Arabidopsis AGL1 e
20	245	27.0	896	25	ABX11510	DNA encoding Arabi
21	245	27.0	1202	21	AAC34634	Arabidopsis thalia
22	236.2	26.0	997	17	AAT34432	Eucalyptus AGE2 cD
23	235.6	25.9	903	21	AAC51039	Arabidopsis thalia
24	232.8	25.6	1143	17	AAT34428	Eucalyptus AGE1 cD
25	230.2	25.3	1057	21	AAC37453	Arabidopsis thalia
26	230	25.3	1053	21	AAC46362	Arabidopsis thalia
27	229.2	25.2	805	24	ABK82084	DNA encoding novel
28	229.2	25.2	1259	24	ABK82131	DNA encoding novel
29	217.8	24.0	1218	21	AAC43678	Zea mays DNA fragm
30	209.8	23.1	581	21	AAC57270	Eucalyptus grandis
31	200.2	22.0	1321	21	AAC47718	Zea mays DNA fragm
32	175.6	19.3	1144	19	AAV18014	Pinus radiata cone
33	175	19.3	557	21	AAC57352	Pinus radiata tran
34	175	19.3	1070	19	AAV18012	Pinus radiata cone
35	172.8	19.0	466	21	AAC35890	Zea mays DNA fragm
36	172.4	19.0	385	21	AAC56258	Pinus radiata tran
37	172.4	19.0	385	21	AAC57315	Pinus radiata tran
38	171.2	18.8	1268	22	AAF76058	Maize MADS-box gen
39	171.2	18.8	1268	22	AAF76445	Maize ZmMADS2 codi
40	171.2	18.8	1280	24	AAD43930	Corn nitrate-respo
41	167.2	18.4	511	21	AAC33755	Arabidopsis thalia
42	165.6	18.2	558	21	AAC56216	Pinus radiata tran
43	164.6	18.1	687	21	AAC42918	Arabidopsis thalia
44	164.6	18.1	1038	24	ABK65164	Arabidopsis cDNA e
45	164.6	18.1	1038	25	AAD47719	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAA99364
ID AAA99364 standard; DNA; 909 BP.
XX
AC AAA99364;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter DNA sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Pinus radiata.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART-) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT-) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;
XX
DR WPI; 2000-594442/56.
XX
PT Novel plant reproductive tissue promoter, useful to produce plants

PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 8; Page 39-40; 51pp; English.
XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 909 BP; 299 A; 172 C; 216 G; 222 T; 0 other;

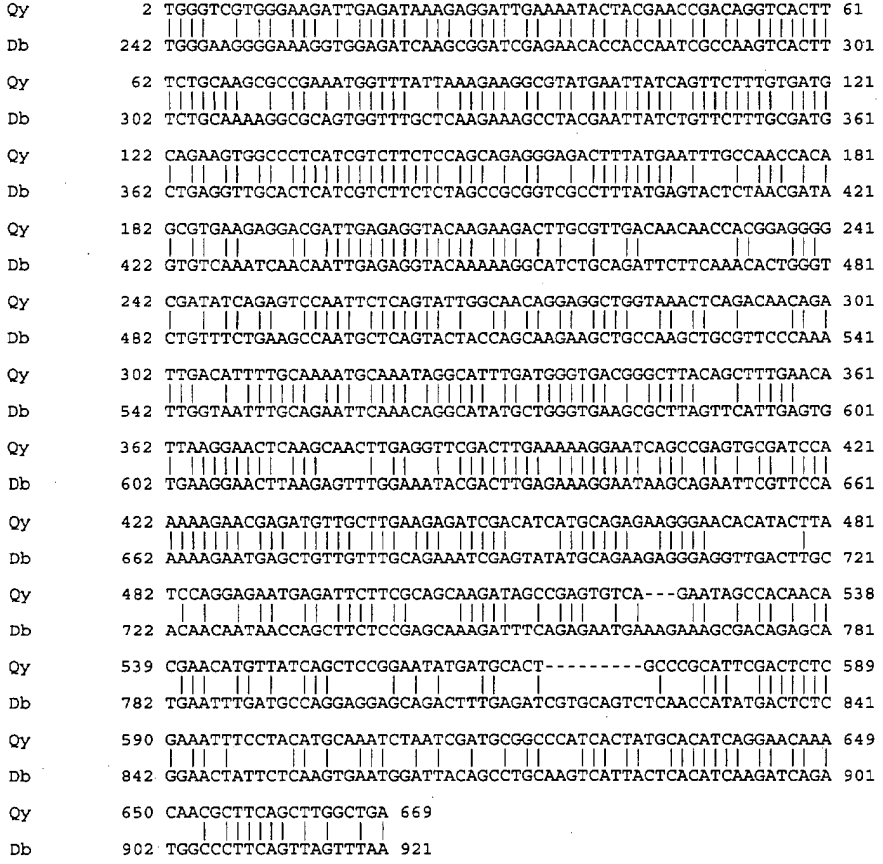
Query Match 100.0%; Score 909; DB 21; Length 909;
Best Local Similarity 100.0%; Pred. No. 3.4e-261;
Matches 909; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Qy        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTGGCTGAACGTTGAAGCGGTGGACGCTTAAAACTCAATCAAGGCACCCGAAAAATATG 720

Qy        721 CTAGTAACCTTGAATGAGATTCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTAGTAACCTTGAATGAGATTCAGAGTCGAAATATTGCGAGGCAAGAGCACAATGGAAGA 780

Qy        781 GATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GATAGCTCCTAGTATGAATATGGATTTATGATATTAACATATGGTTTGTCAGCTTTAAAT 840

Qy        841 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAGCTGTTTGAAACAAAGAATACAACATATTAGCTAGTATTTTTTTGGCGCATGTTATC 900

Qy        901 TTTCTGTTG 909
              |||||||||
Db        901 TTTCTGTTG 909
```

RESULT 2
AAZ57947
ID AAZ57947 standard; cDNA; 1201 BP.
XX
AC AAZ57947;
XX
DT 11-APR-2000 (first entry)
XX
DE Poplar floral homeotic gene PTAG-1 cDNA.
XX
KW Poplar; PTAG-1; floral homeotic gene; transgenic plant; sterility;
KW fertility; ss.
XX
OS Populus balsamifera subsp. trichocarpa.
XX
FH Key Location/Qualifiers
FT CDS 196..921
FT /*tag= a
FT /note= "the coding region is also specifically
FT claimed in Claim 1"
XX
PN CA2227940-A1.
XX
PD 06-OCT-1999.
XX
PF 07-APR-1998; 98CA-2227940.
XX
PR 06-APR-1998; 98US-0080851.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Rottmann WH, Brunner AM, Sheppard LA, Strauss SH;
XX
DR WPI; 2000-106662/10.
DR P-PSDB; AAY58656.
XX
PT Nucleic acid from Populus trichocarpa genes, useful for producing
PT transgenic plants, particularly trees, with modified fertility
PT characteristics such as sterility -
XX
PS Claim 1; Page 64-66; 92pp; English.
XX
CC The present sequence is that of cDNA encoding Populus balsamifera
CC subsp. trichocarpa PTAG-1 protein (see AAY58656). The PTAG-1 gene (see
CC AAZ57946) is 1 of 4 newly identified floral homeotic genes from this
CC poplar species. PTAG-1 is a homologue of AGAMOUS and is expressed
CC in floral tissue. The invention provides nucleic acid sequences of
CC these 4 Populus genes, the corresponding cDNA sequences (see
CC AAZ57942-49) and deduced amino acid sequences (see AAY58454-57). It also
CC provides methods of using the genes and cDNA to produce genetically
CC engineered Populus species and other trees having modified fertility
CC characteristics, including sterility. Genetic constructs useful in
CC producing these genetically engineered trees include antisense
CC versions of PTAG-1, dominant negative mutants, and constructs useful
CC for sense suppression. Promoter sequences may be used to obtain
CC floral specific expression of genes such as cytotoxins that may be
CC used in genetic ablation strategies to produce trees having modified
CC fertility characteristics, including sterility. Sterile trees
CC allow increased wood yield and a reduction in the production of
CC allergens such as pollen.

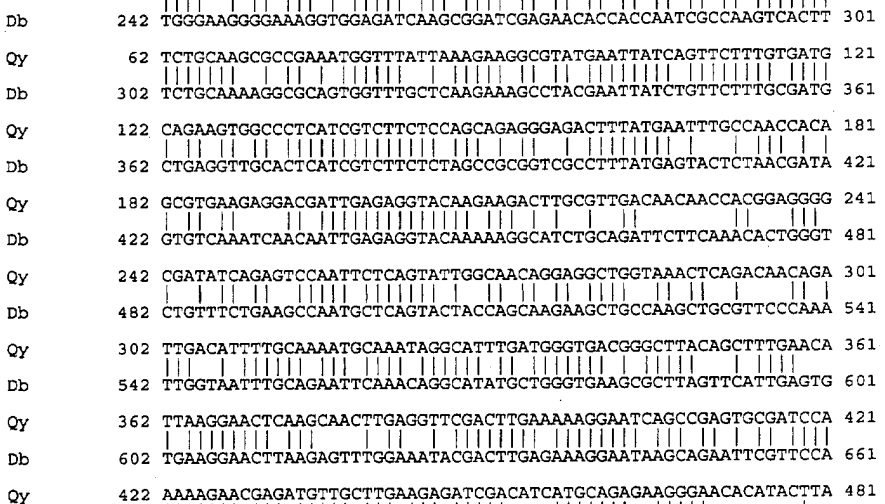
XX
SQ Sequence 1201 BP; 354 A; 251 C; 257 G; 339 T; 0 other;

Query Match 28.6%; Score 260; DB 21; Length 1201;
Best Local Similarity 63.7%; Pred. No. 3.8e-67;
Matches 433; Conservative 0; Mismatches 235; Indels 12; Gaps 2;



RESULT 3
AAF85398
ID AAF85398 standard; cDNA; 1219 BP.
XX
AC AAF85398;
XX
DT 23-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the floral homeotic protein PTAG-1.
XX
KW Floral homeotic gene; PTD; PTLF; PTAG-1; PTAG-2; floral tissue; LEAFY;
KW LFY; FLORICAULA; FLO; DEFICIENS; DEF; AGAMOUS; AG; transgenic plant;
KW fertility; sterility; ss.
XX
OS Populus balsamifera.
XX
FH Key Location/Qualifiers
FT CDS 196..921
FT /*tag= a
FT /product= "PTAG-1"
XX
PN CA2319853-A1.
XX
PD 01-APR-2001.
XX
PF 02-OCT-2000; 2000CA-2319853.
XX
PR 01-OCT-1999; 99US-0410464.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Rottman WH, Strauss SH, Brunner AM, Sheppard LA;
XX
DR WPI; 2001-336098/36.
DR P-PSDB; AAB68437.
XX
PT Novel isolated polynucleotide derived from Populus species, useful for
PT producing transgenic plants having modified fertility characteristic,
PT particularly sterility -
XX
PS Claim 25; Page 54-55; 69pp; English.
XX
CC The present sequence encodes a floral homeotic protein, designated
CC PTAG-1, which is derived from Populus balsamifera subsp. trichocarpa.
CC The specification also describes PTD, PTLF, and PTAG-2 proteins. The
CC floral homeotic proteins are expressed in floral tissues. PTLF is a
CC homologue of LEAFY (LFY) and FLORICAULA (FLO), and is expressed in
CC immature inflorescences on which floral primordia are developing. PTD
CC is a homologue of DEFICIENS (DEF), and is strongly expressed in stamen
CC primordia from the onset of organogenesis. PTAG-1 and PTAG-2 are
CC homologues of AGAMOUS (AG). The floral homeotic proteins and
CC polynucleotides are useful for producing transgenic plants having
CC modified fertility characteristics, particularly sterility.
XX
SQ Sequence 1219 BP; 371 A; 252 C; 257 G; 339 T; 0 other;

Query Match 28.6%; Score 260; DB 22; Length 1219;
Best Local Similarity 63.7%; Pred. No. 3.8e-67;
Matches 433; Conservative 0; Mismatches 235; Indels 12; Gaps 2;



```
Db        662 AAAAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGC 721

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCA---GAATAGCCACAACA 538

Db        722 ACAACAATAACCAGCTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCA 781

Qy        539 CGAACATGTTATCAGCTCCGGAATATGATGCACT---------GCCCGCATTCGACTCTC 589

Db        782 TGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTC 841

Qy        590 GAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAA 649

Db        842 GGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGA 901

Qy        650 CAACGCTTCAGCTTGGCTGA 669

Db        902 TGGCCCTTCAGTTAGTTTAA 921
```

RESULT 4
ABK88489
ID ABK88489 standard; cDNA; 1219 BP.
XX
AC ABK88489;
XX
DT 07-OCT-2002 (first entry)
XX
DE Poplar floral homeotic protein, PTAG-1, cDNA.
XX
KW Poplar; ss; gene; plant; DEFICIENS; transgenic; promoter;
KW protein transduction domain; floral homeotic gene;
KW floral-specific expression; cytotoxin; fertility; sterility;
KW PTLF; PTAG-1; PTAG-2.
XX
OS Populus balsamifera subsp. trichoc arpa.
XX
FH Key   Location/Qualifiers
FT CDS   196..921
FT   /*tag= a
FT   /product= "PTAG-1"
XX
PN US6395892-B1.
XX
PD 28-MAY-2002.
XX
PF 01-OCT-1999; 99US-0410464.
XX
PR 06-APR-1998; 98US-080851P.
PR 06-APR-1999; 99US-0287700.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Strauss SH, Rottmann W, Brunner A, Sheppard L;
XX
DR WPI; 2002-572853/61.
DR P-PSDB; ABG30867.
XX
PT New protein transduction domain promoter nucleic acid molecule useful
PT for producing transgenic plants having modified fertility
PT characteristics, particularly sterility -
XX
PS Disclosure; Column 63-66; 46pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule especially a
CC protein transduction domain (PTD) promoter: (i) that hybridises under
CC wash conditions of 0.2 x SSC (saline sodium citrate), 0.1 %SDS (sodium
CC dodecyl sulphate) at 65 plusoC to nucleotides or (ii) comprising 35
CC consecutive nucleotides of the PTD gene. PTD is a floral homeotic
CC gene and is the homologue of DEFICIENS. Also includes are a recombinant
CC nucleic acid comprising the PTD promoter, a cell transformed with
CC the recombinant nucleic acid and a transgenic plant comprising the
CC transformed cell. The PTD promoter is useful to obtain floral-specific
CC expression of genes such as cytotoxins, that are employed in genetic
CC ablation strategies to produce trees having modified fertility
CC characteristics, including sterility. Genetic constructs comprising
CC antisense versions or dominant negative mutants of PTD are useful in
CC producing genetically engineered Poplars and other trees, and for sense
CC suppression. Also disclosed are 3 other homeotic genes PTLF, PTAG-1
CC and PTAG-2 (none are defined). The present sequence is the PTAG-1 cDNA.
XX
SQ Sequence 1219 BP; 371 A; 252 C; 257 G; 339 T; 0 other;

  Query Match   28.6%; Score 260; DB 24; Length 1219;
  Best Local Similarity 63.7%; Pred. No. 3.8e-67;
  Matches 433; Conservative   0; Mismatches 235; Indels   12; Gaps   2;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61

Db        242 TGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 301

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121

Db        302 TCTGCAAAAGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATG 361

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181

Db        362 CTGAGGTTGCACTCATCGTCTTCTCTAGCCGCGGTCGCCTTTATGAGTACTCTAACGATA 421

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241

Db        422 GTGTCAAATCAACAATTGAGAGGTACAAAAAGGCATCTGCAGATTCTTCAAACACTGGGT 481

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301

Db        482 CTGTTTCTGAAGCCAATGCTCAGTACTACCAGCAAGAAGCTGCCAAGCTGCGTTCCCAAA 541

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361

Db        542 TTGGTAATTTGCAGAATTCAAACAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTG 601

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421

Db        602 TGAAGGAACTTAAGAGTTTGGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCA 661

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481

Db        662 AAAAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGC 721

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCA---GAATAGCCACAACA 538

Db        722 ACAACAATAACCAGCTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCA 781

Qy        539 CGAACATGTTATCAGCTCCGGAATATGATGCACT---------GCCCGCATTCGACTCTC 589

Db        782 TGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTC 841

Qy        590 GAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAA 649

Db        842 GGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGA 901

Qy        650 CAACGCTTCAGCTTGGCTGA 669

Db        902 TGGCCCTTCAGTTAGTTTAA 921
```

RESULT 5
AAF85399
ID AAF85399 standard; cDNA; 723 BP.
XX
AC AAF85399;
XX
DT 23-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the floral homeotic protein PTAG-1.
XX
KW Floral homeotic gene; PTD; PTLF; PTAG-1; PTAG-2; floral tissue; LEAFY;

KW LFY; FLORICAULA; FLO; DEFICIENS; DEF; AGAMOUS; AG; transgenic plant;
KW fertility; sterility; ss.
XX
OS Populus balsamifera.
XX
FH Key Location/Qualifiers
FT CDS 1..723
FT /*tag= a
FT /product= "PTAG-1"
XX
PN CA2319853-A1.
XX
PD 01-APR-2001.
XX
PF 02-OCT-2000; 2000CA-2319853.
XX
PR 01-OCT-1999; 99US-0410464.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Rottman WH, Strauss SH, Brunner AM, Sheppard LA;
XX
DR WPI; 2001-336098/36.
DR P-PSDB; AAB68437.
XX
PT Novel isolated polynucleotide derived from Populus species, useful for
PT producing transgenic plants having modified fertility characteristic,
PT particularly sterility -
XX
PS Claim 25; Page 55-56; 69pp; English.
XX
CC The present sequence encodes a floral homeotic protein, designated
CC PTAG-1, which is derived from Populus balsamifera subsp. trichocarpa.
CC The specification also describes PTD, PTLF, and PTAG-2 proteins. The
CC floral homeotic proteins are expressed in floral tissues. PTLF is a
CC homologue of LEAFY (LFY) and FLORICAULA (FLO), and is expressed in
CC immature inflorescences on which floral primordia are developing. PTD
CC is a homologue of DEFICIENS (DEF), and is strongly expressed in stamen
CC primordia from the onset of organogenesis. PTAG-1 and PTAG-2 are
CC homologues of AGAMOUS (AG). The floral homeotic proteins and
CC polynucleotides are useful for producing transgenic plants having
CC modified fertility characteristics, particularly sterility.
XX
SQ Sequence 723 BP; 226 A; 147 C; 175 G; 175 T; 0 other;

Query Match 28.6%; Score 259.8; DB 22; Length 723;
Best Local Similarity 63.9%; Pred. No. 3.4e-67;
Matches 431; Conservative 0; Mismatches 232; Indels 12; Gaps 2;

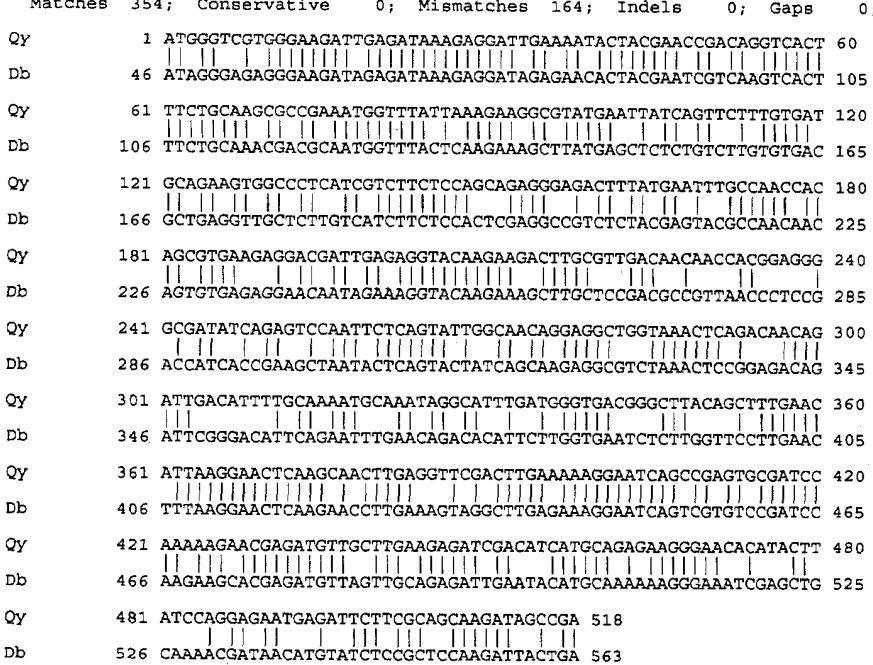
```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              |||| | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db         47 TGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 106

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              ||||||| | || | |||||| | ||||| || || |||||||| |||||||| ||||
Db        107 TCTGCAAAAGGCGCAGTGGTTTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATG 166

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | ||| || || |||||||||||||||| ||| | || | |||||||| | ||| | |
Db        167 CTGAGGTTGCACTCATCGTCTTCTCTAGCCGCGGTCGCCTTTATGAGTACTCTAACGATA 226

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||     || |||||||||||||| ||| | |   |  ||         ||   |||
Db        227 GTGTCAAATCAACAATTGAGAGGTACAAAAAGGCATCTGCAGATTCTTCAAACACTGGGT 286

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  ||||| |||||||| |  || || || ||||  || ||  |    || |
Db        287 CTGTTTCTGAAGCCAATGCTCAGTACTACCAGCAAGAAGCTGCCAAGCTGCGTTCCCAAA 346

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | |||||| ||| |||| |||||| || |||||||  | |||||    | ||||
Db        347 TTGGTAATTTGCAGAATTCAAACAGGCATATGCTGGGTGAAGCGCTTAGTTCATTGAGTG 406

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |||||||| |||     | ||  | |||||||| |||||||| ||| || | || ||||
Db        407 TGAAGGAACTTAAGAGTTTGGAAATACGACTTGAGAAAGGAATAAGCAGAATTCGTTCCA 466

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| ||||| ||| ||| |||||     |||||||  |||||          |
Db        467 AAAAGAATGAGCTGTTGTTTGCAGAAATCGAGTATATGCAGAAGAGGGAGGTTGACTTGC 526

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCA---GAATAGCCACAACA 538
              | | | ||  || |||| ||     |||||  | |||  | |    ||  | || | ||
Db        527 ACAACAATAACCAGCTTCTCCGAGCAAAGATTTCAGAGAATGAAAGAAAGCGACAGAGCA 586

Qy        539 CGAACATGTTATCAGCTCCGGAATATGATGCACT---------GCCCGCATTCGACTCTC 589
              |||  || |  |||       | | |   ||   |         |  |||  |||||||
Db        587 TGAATTTGATGCCAGGAGGAGCAGACTTTGAGATCGTGCAGTCTCAACCATATGACTCTC 646

Qy        590 GAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAA 649
              | || |      || |  ||| | |  | ||  ||| ||  |||||||| || || |
Db        647 GGAACTATTCTCAAGTGAATGGATTACAGCCTGCAAGTCATTACTCACATCAAGATCAGA 706

Qy        650 CAACGCTTCAGCTTG 664
                | |||||| | |
Db        707 TGGCCCTTCAGTTAG 721
```

RESULT 6
AAC43464
ID AAC43464 standard; DNA; 777 BP.
XX
AC AAC43464;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 39336.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

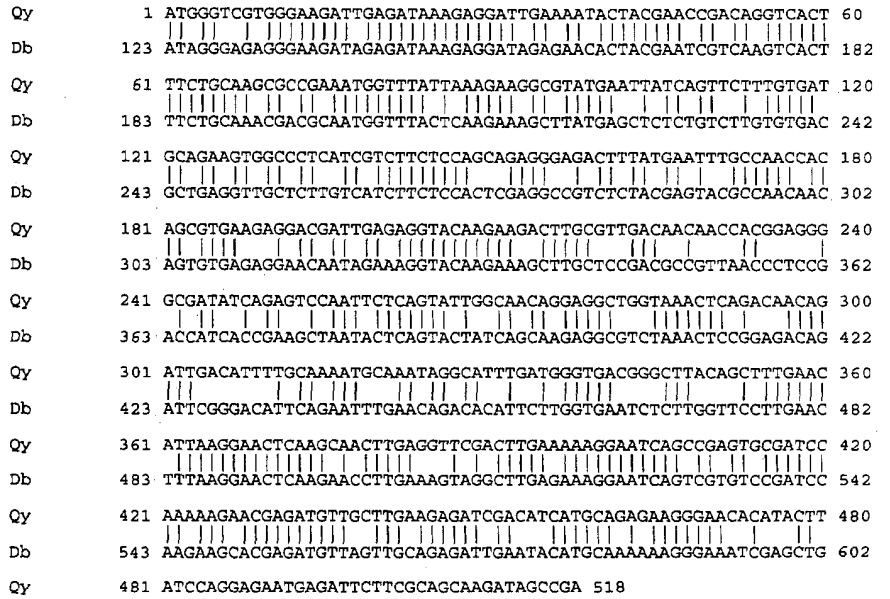
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 28.1%; Score 255.6; DB 21; Length 777;
Best Local Similarity 68.3%; Pred. No. 6.3e-66;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;



RESULT 7
AAV99860
ID AAV99860 standard; cDNA; 959 BP.
XX
AC AAV99860;
XX
DT 10-MAY-1999 (first entry)
XX
DE Arabidopsis AGL5 cDNA.
XX
KW AGL5-like gene; agamous-like 5 gene; seed dispersal; dehiscence;
KW transgenic plant; promoter; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 78..818
FT /*tag= a
XX
PN WO9900502-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-US13208.
XX
PR 28-APR-1998; 98US-0067800.
PR 27-JUN-1997; 97US-0051030.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Ferrandiz C, Yanofsky MF;
XX
DR WPI; 1999-095747/08.
DR P-PSDB; AAW81000.
XX
PT Use of agamous-like nucleic acids - useful for the production of
PT transgenic seed plants in which dehiscence is modified resulting in
PT delayed seed dispersal
XX
PS Disclosure; Page 90-91; 126pp; English.
XX
CC This is the nucleotide sequence of an Arabidopsis agamous-like 5
CC (AGL5) gene cDNA clone that codes for AGL5 polypeptide (see
CC AAW81000) which is involved in the regulation of dehiscence. The
CC invention provides a transgenic seed plant, such as an agl5 and
CC agl1 double mutant, that is characterised by delayed seed dispersal
CC due to suppression of AGL5 and AGL1 (see AAV99859) expression. The
CC invention also provides dehiscence zone regulatory elements, derived
CC from the AGL1 or AGL5 gene (see AAV99857 and AAV99858), that confer
CC selective expression upon an operatively linked nucleic acid
CC molecule in the valve margin or dehiscence zone of a seed plant.
CC The transgenic seed plants include members of the Brassicaceae,
CC such as rapeseed, and members of the Fabaceae, such as soybean,
CC pea, lentil and bean. A plant expression vector comprising a
CC dehiscence zone-selective regulatory element, and a kit for
CC producing the transgenic seed plants are also provided.
XX
SQ Sequence 959 BP; 297 A; 218 C; 214 G; 230 T; 0 other;

Query Match 28.1%; Score 255.6; DB 20; Length 959;
Best Local Similarity 68.3%; Pred. No. 7e-66;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;



```
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640


RESULT 8
ABQ79456
ID   ABQ79456 standard; DNA; 959 BP.
XX
AC   ABQ79456;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Arabidopsis AGL5 encoding sequence.
XX
KW   Plant; lignin; AGL8; lignification; vascular plant; woody plant;
KW   eucalyptus; cottonwood; alder; Douglas fir; hemlock; pine; spruce;
KW   leguminous plant; alfalfa; clover; lucerne; birdsfoot trefoil;
KW   stylosanthes; lotononis bainessii; sainfoin; forage grass; bahiagrass;
KW   bermudagrass; dallisgrass; pangolagrass; big bluestem; Indiangrass;
KW   switchgrass; smooth bromegrass; orchardgrass; timothy;
KW   Kentucky bluegrass; tall fescue; agriculture; wood product; AGL5; gene;
KW   ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             78..818
FT                   /*tag=  a
FT                   /product= "AGL5"
XX
PN   US6410826-B1.
XX
PD   25-JUN-2002.
XX
PF   25-JUN-1999;   99US-0339998.
XX
PR   25-JUN-1998;   98US-090649P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Yanofsky MF,  Liljegren S,  Ferrandiz C;
XX
DR   WPI; 2002-588897/63.
DR   P-PSDB; ABB98194.
XX
PT   Reducing lignification in vascular plants e.g. woody plant, forage
PT   grass, or leguminous plant comprises ectopically expressing a nucleic
PT   acid molecule encoding an Arabidopsis AGL8-like gene product in the
PT   vascular plant  -
XX
PS   Disclosure; Column 41-44; 41pp; English.
XX
CC   The invention relates to reducing lignification in a vascular plant by
CC   introducing an exogenous nucleic acid molecule encoding an AGL8-like
CC   gene product (a MADS box protein). This results in lignification being
CC   reduced due to ectopic expression of the nucleic acid molecule. The
CC   method of the invention is useful for reducing lignification in a
CC   vascular plant, including a woody plant, such as Eucalyptus, cottonwood,
CC   alder, Douglas fir, Hemlock, pine and spruce, a leguminous plant such as
CC   alfalfa, clover, lucerne, birdsfoot trefoil, Stylosanthes, Lotononis
CC   bainessii and sainfoin, or a forage grass such as bahiagrass,
CC   bermudagrass, dallisgrass, pangolagrass, big bluestem, Indiangrass,
CC   switchgrass, smooth bromegrass, orchardgrass, timothy, Kentucky bluegrass
CC   and tall fescue. The method is useful for developing plant varieties
CC   with improved characteristics. This is useful in agriculture and the wood
CC   products industry, both benefiting by a reduction in lignin content. The
CC   current sequence represents the Arabidopsis AGL5 encoding sequence.
XX
SQ   Sequence 959 BP; 297 A; 218 C; 214 G; 230 T; 0 other;

  Query Match             28.1%;  Score 255.6;  DB 24;  Length 959;
  Best Local Similarity   68.3%;  Pred. No. 7e-66;
  Matches  354;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              || ||   | ||||||||| |||||||||||||||| || || |||||||| || || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || ||||||||| | ||||| || ||||| | || || | |||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || || ||  || ||||||||||   |||| | || || || | |||||| ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || ||||   | || || || ||||||||||| |||||  |||  |  ||    |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | || | || | ||| ||||||| | || || |||||  |||||||| |  ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||       | || |||  || || ||  |  | |||||    |||    | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | |||||   | | ||||| |||||||||| || || ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| |||||||||| ||| |||||| ||   |||||| | |||||||  |  ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                | ||  ||   |  ||| |||  |||||  | ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640


RESULT 9
ABX11511
ID   ABX11511 standard; DNA; 959 BP.
XX
AC   ABX11511;
XX
DT   01-MAY-2003  (first entry)
XX
DE   DNA encoding Arabidopsis Agamous-like protein 5 (AGL5).
XX
KW   Agamous-like protein; AGL1; AGL1-like protein; vascular plant;
KW   woody plant; leguminous plant; forage grass; lignification; tall fescue;
KW   eucalyptus; bahiagrass; bermudagrass; dallisgrass; smooth bromegrass;
KW   pangolagrass; big bluestem; indiangrass; switchgrass; alfalfa;
KW   orchardgrass; timothy grass; Kentucky bluegrass; animal husbandry;
KW   gene; ds; transgenic.
XX
OS   Arabidopsis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             78..818
FT                   /*tag=  a
FT                   /product= "AGL5"
FT                   /note= "Agamous-like protein 5"
XX
PN   US2002129403-A1.
XX
PD   12-SEP-2002.
XX
PF   15-OCT-2001; 2001US-0978730.
XX
PR   25-JUN-1998;   98US-090649P.
PR   25-JUN-1999;   99US-0339998.
XX
PA   (REGC ) UNIV CALIFORNIA.
```

```
XX
PI   Yanofsky MF,  Liljegren S;
XX
DR   WPI; 2003-147404/14.
DR   P-PSDB; ABG75909.
XX
PT   Vascular plant lignification reduction method for genetic engineering,
PT   involves expressing nucleic acid molecule encoding specific type of
PT   gene product in vascular plant, ectopically  -
XX
PS   Disclosure; Fig 8; 46pp; English.
XX
CC   The invention describes a nucleic acid molecule encoding an AGL8-like
CC   gene product comprising a amino acid sequences of an AGL8 orthologue
CC   such as Arabidopsis AGL8 expressed ectopically in a vascular plant such
CC   as woody plant, leguminous plant or forage grass. The nucleic acids are
CC   useful for reducing lignification in woody plants such as eucalyptus,
CC   leguminous forage crops such as alfalfa, and in forage grasses such as
CC   bahiagrass, bermudagrass, dallisgrass, pangolagrass, big bluestem,
CC   indiangrass, switchgrass, smooth bromegrass, orchardgrass, timothy,
CC   Kentucky bluegrass or tall fescue used for plant/agricultural genetic
CC   engineering. Modification of lignification allows more efficient use of
CC   plant biomass in animal husbandry, by enabling production of
CC   lignification enhanced/reduced vascular plants,and enables lignification
CC   changes in monocotyledonous or dicotyledonous angiosperm or with a
CC   gymnosperm, which is a seed-bearing plant with seeds not enclosed in
CC   ovary. This sequence encodes Arabidopsis Agamous-like protein 5 (AGL5),
CC   a protein that regulates the process of lignification.
XX
SQ   Sequence 959 BP; 297 A; 218 C; 214 G; 230 T; 0 other;

  Query Match             28.1%;  Score 255.6;  DB 25;  Length 959;
  Best Local Similarity   68.3%;  Pred. No. 7e-66;
  Matches  354;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||  ||   | |||||||| |||||||||||||||| || || |||||||| || || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || |||||||||| | ||||| || |||||   | ||  ||  | ||||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || || ||  || |||||||||    ||||  | || || || |  |||||| ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || ||||   | || || || ||||||||||  |||||  |||  |   ||      |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||  | ||| |||||||  |  || || |||||   |||||| |    ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||       | || |||   || || ||  |  | |||||     |||    | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | |||||    | |||||  |||||||||| || || ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| |||||||||  ||| |||||| ||   |||||| | |||||||   |   ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                | || ||   |  ||| |||  ||||||  | ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640

RESULT 10
```

```
AAZ50201
ID   AAZ50201 standard; cDNA; 1157 BP.
XX
AC   AAZ50201;
XX
DT   17-MAY-2000  (first entry)
XX
DE   Petunia hybrida nectary-specific protein FBP15 cDNA.
XX
KW   FBP15; nectary-specific protein; metabolite; recombinant protein;
KW   transgenic plant; nectar; honey; insect; bee; pharmaceutical; enzyme;
KW   biotest; antioxidant; food additive; MADS-box protein; ss.
XX
OS   Petunia hybrida.
XX
FH   Key             Location/Qualifiers
FT   CDS             179..844
FT                   /*tag=  a
FT                   /product= "FBP15 protein"
FT   mat_peptide     206..361
FT                   /*tag=  b
FT                   /product= "MADS-box"
FT                   /note= "homologous to a similar region in the Arabidopsis
FT                   agamous protein and Antirrhinum deficiens protein"
FT   misc_feature    503..616
FT                   /*tag=  c
FT                   /label= "K-box"
XX
PN   WO200004176-A1.
XX
PD   27-JAN-2000.
XX
PF   15-JUL-1999;   99WO-NL00453.
XX
PR   16-JUL-1998;   98EP-0202375.
PR   14-DEC-1998;   98EP-0204215.
XX
PA   (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.
XX
PI   Creemers J,  Angenent GC,  Kater MM;
XX
DR   WPI; 2000-182438/16.
DR   P-PSDB; AAY44804.
XX
PT   New DNA encoding Petunia hybrida nectary-specific proteins, useful for,
PT   e.g. producing modified honey  -
XX
PS   Claim 12; Page 43; 93pp; English.
XX
CC   The present sequence is a cDNA encoding nectary-specific protein FBP15.
CC   It was isolated from a cDNA library made from nectaries of Petunia
CC   hybrida strain W115 flowers. FBP15 is a MADS-box protein which is
CC   exclusively expressed in the nectaries of P. hybrida.
CC   A DNA sequence from the promoter region upstream of
CC   nectary-specific expressed sequence e.g. NEC1 and FBP15 DNAs is used in
CC   a recombinant DNA construct comprising a DNA encoding a metabolite
CC   preferably recombinant protein, a DNA encoding a signal peptide that
CC   targets the recombinant protein to the nectar and optionally a signal
CC   sequence functional in plants for the transcription termination and
CC   polyadenylation of an RNA molecule. The DNA construct is useful for
CC   producing transgenic plants which excrete recombinant proteins in its
CC   nectar. The nectar is processed into honey by insects (preferably bees)
CC   and the desired protein is easily recovered from it. The recombinant
CC   proteins are useful for pharmaceutical purposes, as enzymes for biotests
CC   and antioxidants for food additives.
XX
SQ   Sequence 1157 BP; 391 A; 203 C; 229 G; 334 T; 0 other;

  Query Match             27.7%;  Score 251.4;  DB 21;  Length 1157;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-64;
  Matches  415;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
```

```
Db        179 ATGGGAAGAGGAAAGATTGAGATTAAGAGGATTGAAAATACAACAAATCGTCAAGTCACT 238

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        239 TTCTGTAAGAGAAGAAATGGGTTGCTTAAAAAAGCTTATGAACTTTCTGTTCTTTGTGAT 298

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        299 GCTGAAGTTGCTCTCATCGTTTTCTCAAGCCGTGGCCGCCTCTATGAATATGCTAACAAC 358

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        359 AGTGTGAAGGCAACAATTGATAGATATAAGAAAGCATCCTCAGATTCCTCCAACACTGGA 418

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        419 TCTACTTCTGAAGCTAACACTCAGTTTTATCAACAAGAAGCTGCCAAACTCCGAGTTCAG 478

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        479 ATTGGTAACTTACAGAACTCAAACAGGAACATGCTAGGCGAGTCTCTAAGTTCTCTGACT 538

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        539 GCAAAAGATCTGAAAGGCCTGGAGACCAAACTTGAGAAAGGAATTAGTAGAATTAGGTCC 598

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        599 AAAAAGAATGAACTCCTGTTTGCTGAGATTGAGTATATGCGAAAAAGGGAAATTGATTTG 658

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        659 CACAACAACAATCAGATGCTTCGGGCAAAGATAGCTGAGAGTGAAAGAAATGTGAACATG 718

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        719 ---ATGGGAGGAGAATTTGAGCTGATGCAATCTCATCCGTACGATCCAAGAGACTTCTTC 775

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Db        776 CAAGTGAACGGCTTACAGCATAATCATCAATATCCACGCCAAGACAACATGGCTCTTCAA 835

Qy        661 CTTGGCTGA 669

Db        836 TTAGTATAA 844
```

RESULT 11
AAZ35494
ID AAZ35494 standard; cDNA; 1157 BP.
XX
AC AAZ35494;
XX
DT 11-APR-2000 (first entry)
XX
DE Petunia *nectary*-specific FBP15 cDNA.
XX
KW FBP14; nectary; nectar; transgenic plant; honey; ds.
XX
OS Petunia hybrida.
XX
FH Key   Location/Qualifiers
FT CDS   179..841
FT   /*tag= a
XX
PN EP974667-A1.
XX
PD 26-JAN-2000.
XX
PF 16-JUL-1998; 98EP-0202375.
XX
PR 16-JUL-1998; 98EP-0202375.
XX
PA (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.
XX
PI Creemers J, Angenent GC, Kater MM;
XX
DR WPI; 2000-108400/10.
DR P-PSDB; AAY58648.
XX
PT Novel DNA sequences used to produce modified honey, the metabolites of
PT which can be isolated and purified -
XX
PS Claim 5; Page 17; 56pp; English.
XX
CC The present sequence is that of Petunia hybrida strain W115 FBP15
CC cDNA that was identified in nectary gland cDNA by differential
CC *expression*. The cDNA encodes a 221-amino acid MADS box protein
CC (see AAY58648) that is specifically expressed in the nectaries of
CC petunia. The present invention provides a method for producing
CC recombinant proteins in honey. The honey is manufactured by *insects*,
CC preferably honeybees, that collect the nectar of transgenic plants.
CC The FBP15 gene and its promoter can be utilised in expression
CC cassettes for the production of transgenic plants that produce a
CC protein of interest in their nectar.
XX
SQ Sequence 1157 BP; 391 A; 203 C; 229 G; 334 T; 0 other;

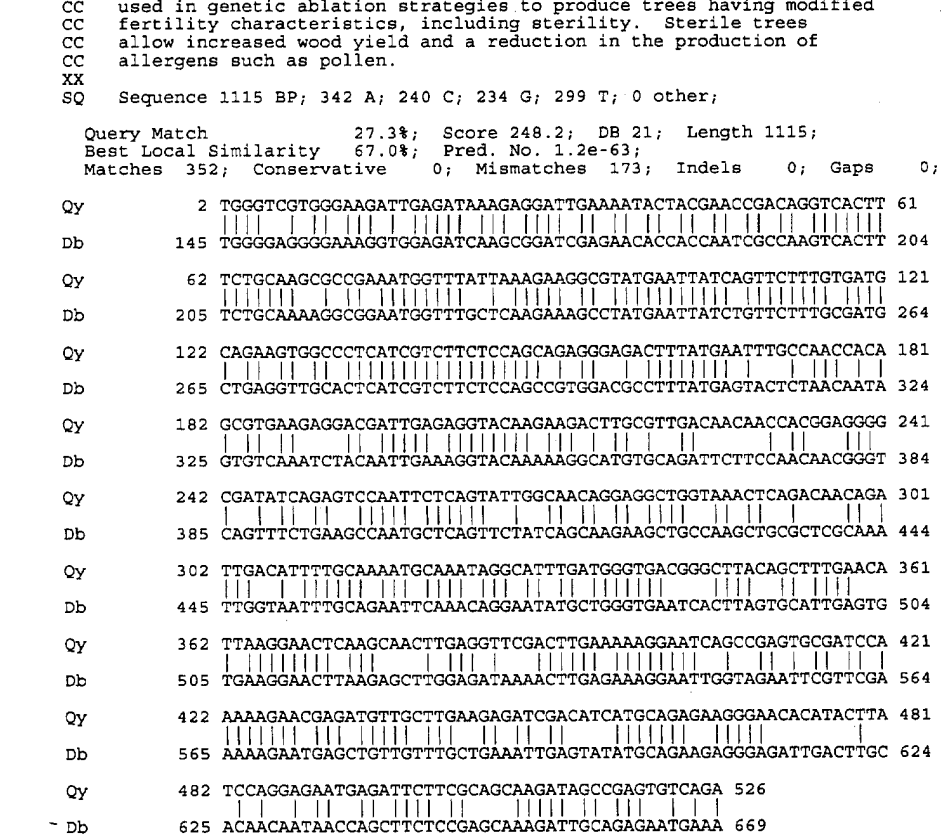
Query Match 27.7%; Score 251.4; DB 21; Length 1157;
Best Local Similarity 62.0%; Pred. No. 1.4e-64;
Matches 415; Conservative 0; Mismatches 251; Indels 3; Gaps 1;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db        179 ATGGGAAGAGGAAAGATTGAGATTAAGAGGATTGAAAATACAACAAATCGTCAAGTCACT 238

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        239 TTCTGTAAGAGAAGAAATGGGTTGCTTAAAAAAGCTTATGAACTTTCTGTTCTTTGTGAT 298

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        299 GCTGAAGTTGCTCTCATCGTTTTCTCAAGCCGTGGCCGCCTCTATGAATATGCTAACAAC 358

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        359 AGTGTGAAGGCAACAATTGATAGATATAAGAAAGCATCCTCAGATTCCTCCAACACTGGA 418

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        419 TCTACTTCTGAAGCTAACACTCAGTTTTATCAACAAGAAGCTGCCAAACTCCGAGTTCAG 478

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        479 ATTGGTAACTTACAGAACTCAAACAGGAACATGCTAGGCGAGTCTCTAAGTTCTCTGACT 538

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        539 GCAAAAGATCTGAAAGGCCTGGAGACCAAACTTGAGAAAGGAATTAGTAGAATTAGGTCC 598

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        599 AAAAAGAATGAACTCCTGTTTGCTGAGATTGAGTATATGCGAAAAAGGGAAATTGATTTG 658

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        659 CACAACAACAATCAGATGCTTCGGGCAAAGATAGCTGAGAGTGAAAGAAATGTGAACATG 718

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        719 ---ATGGGAGGAGAATTTGAGCTGATGCAATCTCATCCGTACGATCCAAGAGACTTCTTC 775

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660

Db        776 CAAGTGAACGGCTTACAGCATAATCATCAATATCCACGCCAAGACAACATGGCTCTTCAA 835
```

```
Qy        661 CTTGGCTGA 669
              | | | ||
Db        836 TTAGTATAA 844


RESULT 12
AAF85402
ID   AAF85402 standard; cDNA; 714 BP.
XX
AC   AAF85402;
XX
DT   23-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the floral homeotic protein PTAG-2.
XX
KW   Floral homeotic gene; PTD; PTLF; PTAG-1; PTAG-2; floral tissue; LEAFY;
KW   LFY; FLORICAULA; FLO; DEFICIENS; DEF; AGAMOUS; AG; transgenic plant;
KW   fertility; sterility; ss.
XX
OS   Populus balsamifera.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..714
FT                   /*tag=  a
FT                   /product= "PTAG-2"
XX
PN   CA2319853-A1.
XX
PD   01-APR-2001.
XX
PF   02-OCT-2000; 2000CA-2319853.
XX
PR   01-OCT-1999;   99US-0410464.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Rottman WH,  Strauss SH,  Brunner AM,  Sheppard LA;
XX
DR   WPI; 2001-336098/36.
DR   P-PSDB; AAB68438.
XX
PT   Novel isolated polynucleotide derived from Populus species, useful for
PT   producing transgenic plants having modified fertility characteristic,
PT   particularly sterility  -
XX
PS   Claim 25; Page 61-62; 69pp; English.
XX
CC   The present sequence encodes a floral homeotic protein, designated
CC   PTAG-2, which is derived from Populus balsamifera subsp. trichocarpa.
CC   The specification also describes PTD, PTLF, and PTAG-1 proteins. The
CC   floral homeotic proteins are expressed in floral tissues. PTLF is a
CC   homologue of LEAFY (LFY) and FLORICAULA (FLO), and is expressed in
CC   immature inflorescences on which floral primordia are developing. PTD
CC   is a homologue of DEFICIENS (DEF), and is strongly expressed in stamen
CC   primordia from the onset of organogenesis. PTAG-1 and PTAG-2 are
CC   homologues of AGAMOUS (AG). The floral homeotic proteins and
CC   polynucleotides are useful for producing transgenic plants having
CC   modified fertility characteristics, particularly sterility.
XX
SQ   Sequence 714 BP; 223 A; 146 C; 172 G; 173 T; 0 other;

  Query Match             27.3%;  Score 248.2;  DB 22;  Length 714;
  Best Local Similarity   67.0%;  Pred. No. 9.9e-64;
  Matches  352;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db         47 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 106

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || |||||||| | ||||| || |||||||||||| |||||||| ||||
Db        107 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 166

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || || |||||||||||||||||| | ||  | |||||||| |  | ||| | |
Db        167 CTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAATA 226

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | ||| ||    || ||||| |||||||| ||| | || |  ||      | ||   |||
Db        227 GTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGGGT 286

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              |  | || ||  |||| |||||| |  || || || |||  || ||  |    || |
Db        287 CAGTTTCTGAAGCCAATGCTCAGTTCTATCAGCAAGAAGCTGCCAAGCTGCGCTCGCAAA 346

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | |||||| ||| |||| ||| || || ||||||||    ||||  || ||||
Db        347 TTGGTAATTTGCAGAATTCAAACAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTG 406

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |||||||| |||   | ||| |  ||||| |||||||  |  || | || ||| |
Db        407 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 466

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| ||||| |||  || || ||    |||||||  |||||          |
Db        467 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 526

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526
              | |  | ||  || |||| ||    ||||  || |||  | | |
Db        527 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 571


RESULT 13
AAZ57949
ID   AAZ57949 standard; cDNA; 1115 BP.
XX
AC   AAZ57949;
XX
DT   11-APR-2000  (first entry)
XX
DE   Poplar floral homeotic gene PTAG-2 cDNA.
XX
KW   Poplar; PTAG-2; floral homeotic gene; transgenic plant; sterility;
KW   fertility; ss.
XX
OS   Populus balsamifera subsp. trichocarpa.
XX
FH   Key             Location/Qualifiers
FT   CDS             99..815
FT                   /*tag=  a
FT                   /note= "the coding region is also specifically
FT                          claimed in Claim 1"
XX
PN   CA2227940-A1.
XX
PD   06-OCT-1999.
XX
PF   07-APR-1998;   98CA-2227940.
XX
PR   06-APR-1998;   98US-0080851.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Rottmann WH,  Brunner AM,  Sheppard LA,  Strauss SH;
XX
DR   WPI; 2000-106662/10.
DR   P-PSDB; AAY58657.
XX
PT   Nucleic acid from Populus trichocarpa genes, useful for producing
PT   transgenic plants, particularly trees, with modified fertility
PT   characteristics such as sterility -
XX
PS   Claim 1; Page 74-75; 92pp; English.
XX
CC   The present sequence is that of cDNA encoding Populus balsamifera
```

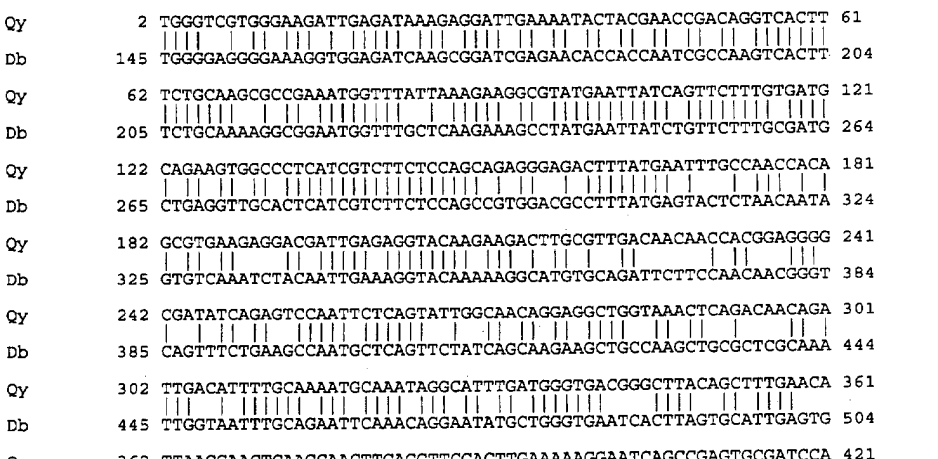
CC subsp. trichocarpa PTAG-2 protein (see AAY58657). The PTAG-2 gene (see
CC AAZ57948) is 1 of 4 newly identified floral homeotic genes from this
CC poplar species. PTAG-2 is a homologue of AGAMOUS and is expressed
CC in floral tissue. The invention provides nucleic acid sequences of
CC these 4 Populus genes, the corresponding cDNA sequences (see
CC AAZ57942-49) and deduced amino acid sequences (see AAY58454-57). It also
CC provides methods of using the genes and cDNA to produce genetically
CC engineered Populus species and other trees having modified fertility
CC characteristics, including sterility. Genetic constructs useful in
CC producing these genetically engineered trees include antisense
CC versions of PTAG-2, dominant negative mutants, and constructs useful
CC for sense suppression. Promoter sequences may be used to obtain
CC floral specific expression of genes such as cytotoxins that may be
CC used in genetic ablation strategies to produce trees having modified
CC fertility characteristics, including sterility. Sterile trees
CC allow increased wood yield and a reduction in the production of
CC allergens such as pollen.
XX
SQ Sequence 1115 BP; 342 A; 240 C; 234 G; 299 T; 0 other;

Query Match 27.3%; Score 248.2; DB 21; Length 1115;
Best Local Similarity 67.0%; Pred. No. 1.2e-63;
Matches 352; Conservative 0; Mismatches 173; Indels 0; Gaps 0;



RESULT 14
AAF85401
ID AAF85401 standard; cDNA; 1159 BP.
XX
AC AAF85401;
XX
DT 23-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the floral homeotic protein PTAG-2.
XX
KW Floral homeotic gene; PTD; PTLF; PTAG-1; PTAG-2; floral tissue; LEAFY;
KW LFY; FLORICAULA; FLO; DEFICIENS; DEF; AGAMOUS; AG; transgenic plant;
KW fertility; sterility; ss.
XX
OS Populus balsamifera.
XX
FH Key Location/Qualifiers
FT CDS 99..815
FT /*tag= a
FT /product= "PTAG-2"
XX
PN CA2319853-A1.
XX
PD 01-APR-2001.
XX
PF 02-OCT-2000; 2000CA-2319853.
XX
PR 01-OCT-1999; 99US-0410464.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Rottman WH, Strauss SH, Brunner AM, Sheppard LA;
XX
DR WPI; 2001-336098/36.
DR P-PSDB; AAB68438.
XX
PT Novel isolated polynucleotide derived from Populus species, useful for
PT producing transgenic plants having modified fertility characteristic,
PT particularly sterility -
XX
PS Claim 25; Page 60-61; 69pp; English.
XX
CC The present sequence encodes a floral homeotic protein, designated
CC PTAG-2, which is derived from Populus balsamifera subsp. trichocarpa.
CC The specification also describes PTD, PTLF, and PTAG-1 proteins. The
CC floral homeotic proteins are expressed in floral tissues. PTLF is a
CC homologue of LEAFY (LFY) and FLORICAULA (FLO), and is expressed in
CC immature inflorescences on which floral primordia are developing. PTD
CC is a homologue of DEFICIENS (DEF), and is strongly expressed in stamen
CC primordia from the onset of organogenesis. PTAG-1 and PTAG-2 are
CC homologues of AGAMOUS (AG). The floral homeotic proteins and
CC polynucleotides are useful for producing transgenic plants having
CC modified fertility characteristics, particularly sterility.
XX
SQ Sequence 1159 BP; 386 A; 240 C; 234 G; 299 T; 0 other;

Query Match 27.3%; Score 248.2; DB 22; Length 1159;
Best Local Similarity 67.0%; Pred. No. 1.2e-63;
Matches 352; Conservative 0; Mismatches 173; Indels 0; Gaps 0;



```
Db        505 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 564

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481

Db        565 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 624

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526

Db        625 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 669
```

RESULT 15
ABK88491
ID ABK88491 standard; cDNA; 1159 BP.
XX
AC ABK88491;
XX
DT 07-OCT-2002 (first entry)
XX
DE Poplar floral homeotic protein, PTAG-2, cDNA.
XX
KW Poplar; ss; gene; plant; DEFICIENS; transgenic; promoter;
KW protein transduction domain; floral homeotic gene;
KW floral-specific expression; cytotoxin; fertility; sterility;
KW PTLF; PTAG-1; PTAG-2.
XX
OS Populus balsamifera subsp. trichoc arpa.
XX
FH Key   Location/Qualifiers
FT CDS   99..815
FT   /*tag= a
FT   /product= "PTAG-2"
XX
PN US6395892-B1.
XX
PD 28-MAY-2002.
XX
PF 01-OCT-1999; 99US-0410464.
XX
PR 06-APR-1998; 98US-080851P.
PR 06-APR-1999; 99US-0287700.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Strauss SH, Rottmann W, Brunner A, Sheppard L;
XX
DR WPI; 2002-572853/61.
DR P-PSDB; ABG30868.
XX
PT New protein transduction domain promoter nucleic acid molecule useful
PT for producing transgenic plants having modified fertility
PT characteristics, particularly sterility -
XX
PS Disclosure; Column 77-80; 46pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule especially a
CC protein transduction domain (PTD) promoter: (i) that hybridises under
CC wash conditions of 0.2 x SSC (saline sodium citrate), 0.1 %SDS (sodium
CC dodecyl sulphate) at 65 plusoC to nucleotides or (ii) comprising 35
CC consecutive nucleotides of the PTD gene. PTD is a floral homeotic
CC gene and is the homologue of DEFICIENS. Also includes are a recombinant
CC nucleic acid comprising the PTD promoter, a cell transformed with
CC the recombinant nucleic acid and a transgenic plant comprising the
CC transformed cell. The PTD promoter is useful to obtain floral-specific
CC expression of genes such as cytotoxins, that are employed in genetic
CC ablation strategies to produce trees having modified fertility
CC characteristics, including sterility. Genetic constructs comprising
CC antisense versions or dominant negative mutants of PTD are useful in
CC producing genetically engineered Poplars and other trees, and for sense
CC suppression. Also disclosed are 3 other homeotic genes PTLF, PTAG-1
CC and PTAG-2 (none are defined). The present sequence is the PTAG-2 cDNA.
XX

SQ Sequence 1159 BP; 386 A; 240 C; 234 G; 299 T; 0 other;

```
  Query Match             27.3%;  Score 248.2;  DB 24;  Length 1159;
  Best Local Similarity   67.0%;  Pred. No. 1.2e-63;
  Matches  352;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61

Db        145 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 204

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121

Db        205 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 264

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181

Db        265 CTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAATA 324

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241

Db        325 GTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGGGT 384

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301

Db        385 CAGTTTCTGAAGCCAATGCTCAGTTCTATCAGCAAGAAGCTGCCAAGCTGCGCTCGCAAA 444

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361

Db        445 TTGGTAATTTGCAGAATTCAAACAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTG 504

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421

Db        505 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 564

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481

Db        565 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 624

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526

Db        625 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 669
```

Search completed: January 21, 2004, 17:31:51
Job time : 238.397 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 3690.68 Seconds
(without alignments)
11217.584 Million cell updates/sec

Title: US-09-936-869-5
Perfect score: 1012
Sequence: 1 atcgaattaaagtcaatggc.....ttctatggtaaaaaaaaaaa 1012

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1012	100.0	1012	8	ATHRNS2X	M98336 Arabidopsis
2		324	32.0	1208	8	AB032256	AB032256 Nicotiana
3		294.4	29.1	1045	8	AMO315591	AJ315591 Antirrhin
4		261.2	25.8	92624	8	AC003000	AC003000 Arabidops
5		249	24.6	9423	8	ATH275979	AJ275979 Arabidops
6		182.8	18.1	981	8	AF139660	AF139660 Calystegi
7		102.2	10.1	4210	8	AB112027	AB112027 Nicotiana
8		82.6	8.2	105384	8	AP006430	AP006430 Lotus jap
9		78.4	7.7	4937	8	AMA489249	AJ489249 Antirrhin
10		55.2	5.5	56284	8	AC007195	AC007195 Arabidops
11	c	54.6	5.4	142955	8	AP003349	AP003349 Oryza sat
12	c	54.6	5.4	145265	8	AP003418	AP003418 Oryza sat
13		51.4	5.1	78756	2	PFMAL6P1_13	Continuation (14 o
14		51.4	5.1	110000	2	PFMAL6P1_12	Continuation (13 o
15		50.8	5.0	103098	8	ATT5C2	AL138664 Arabidops
16		50.4	5.0	1030	8	AY086747	AY086747 Arabidops
17		50.4	5.0	2000	6	AX655393	AX655393 Sequence
18	c	49.2	4.9	188177	2	BX119918	BX119918 Danio rer
19		47.2	4.7	158475	9	AC097463	AC097463 Homo sapi
20		46.8	4.6	693	6	AX412377	AX412377 Sequence
21		46.8	4.6	693	6	AX412567	AX412567 Sequence
22		46.8	4.6	693	6	AX507176	AX507176 Sequence
23		46.8	4.6	763	8	AY114710	AY114710 Arabidops
24		46.8	4.6	881	8	AY072413	AY072413 Arabidops
25		46.8	4.6	925	8	ATU05206	U05206 Arabidopsis
26		46.8	4.6	1141	6	AX083744	AX083744 Sequence
27	c	46.8	4.6	223954	2	AC106472	AC106472 Rattus no
28	c	46.8	4.6	234320	2	AC120569	AC120569 Rattus no
29	c	46.8	4.6	261007	2	AC133419	AC133419 Rattus no
30		46.2	4.6	234894	2	AC111286	AC111286 Rattus no
31		46.2	4.6	265793	2	AC115477	AC115477 Rattus no
32		45.6	4.5	273275	3	AE014828	AE014828 Plasmodiu
33		45.4	4.5	935	8	CAR012689	AJ012689 Cicer ari
34		45.2	4.5	177363	9	AC073856	AC073856 Homo sapi
35		45.2	4.5	201359	2	AC128640	AC128640 Homo sapi
36	c	45	4.4	233746	2	AC120480	AC120480 Rattus no
37		45	4.4	270166	2	AC108263	AC108263 Rattus no
38	c	45	4.4	270972	2	AC132556	AC132556 Rattus no
39		44.8	4.4	331	8	AF504288	AF504288 Crataegus
40		44.6	4.4	506	8	AY260449	AY260449 Prunus ar
41		44.4	4.4	505	6	BD119695	BD119695 EST and e
42	c	44.2	4.4	125020	9	AF429315	AF429315 Homo sapi
43	c	44.2	4.4	192581	2	PFMAL13P1	AL049180 Plasmodiu
44		44	4.3	894	8	LERNALE	X79337 L.esculentu
45	c	43.6	4.3	21024	1	AE014269	AE014269 Streptoco

ALIGNMENTS

RESULT 1
ATHRNS2X
LOCUS ATHRNS2X 1012 bp mRNA linear PLN 30-OCT-1994
DEFINITION Arabidopsis thaliana ribonuclease (RNS2) mRNA, complete cds.
ACCESSION M98336
VERSION M98336.1 GI:289209
KEYWORDS ribonuclease.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1012)
AUTHORS Taylor,C.B., Bariola,P.A., delCardayre,S.B., Raines,R.T. and Green,P.J.

```
  TITLE     RNS2: a senescence-associated RNase of Arabidopsis that diverged
            from the S-RNases before speciation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (11), 5118-5122 (1993)
  MEDLINE   93281708
   PUBMED   8506358
COMMENT     Original source text: Arabidopsis thaliana cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
     gene            1. .1012
                     /gene="RNS2"
     CDS             16. .795
                     /gene="RNS2"
                     /codon_start=1
                     /product="ribonuclease"
                     /protein_id="AAA51406.1"
                     /db_xref="GI:289210"
                     /translation="MASRLCLLLLVACIAGAFAGDVIELNRSQREFDYFALSLQWPGT
                     YCRGTRHCCSKNACCRGSDAPTQFTIHGLWPDYNDGSWPSCCYRSDFKEKEISTLMDG
                     LEKYWPSLSCGSPSSCNGGKGSFWGHEWEKHGTCSSPVFHDEYNYFLTTLNLYLKHNV
                     TDVLYQAGYVASNSEKYPLGGIVTAIQNAFHITPEVVCKRDAIDEIRICFYKDFKPRD
                     CVGSQDLTSRKSCPKYVSLPEYTPLDGEAMVLKMPTEREAL"
BASE COUNT      287 a    197 c    217 g    311 t
ORIGIN
```

```
  Query Match           100.0%;  Score 1012;  DB 8;  Length 1012;
  Best Local Similarity  100.0%;  Pred. No. 6.8e-247;
  Matches 1012;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60

Qy         61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120

Qy        121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180

Qy        181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240

Qy        241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300

Qy        301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360

Qy        361 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 420

Qy        421 TGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTAC 480

Qy        481 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 540

Qy        541 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 600

Qy        601 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660

Qy        661 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 720

Qy        721 AGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 780

Qy        781 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 840

Qy        841 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 900

Qy        901 TATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960

Qy        961 AGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
```

```
RESULT 2
AB032256
LOCUS       AB032256               1208 bp    mRNA    linear   PLN 17-SEP-1999
DEFINITION  Nicotiana glutinosa NGR2 mRNA for RNase NGR2, complete cds.
ACCESSION   AB032256
VERSION     AB032256.1  GI:5902453
KEYWORDS    RNase NGR2.
SOURCE      Nicotiana glutinosa
  ORGANISM  Nicotiana glutinosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (sites)
  AUTHORS   Kariu,T., Kurata,N., Yamasaki,N. and Kimura,M.
  TITLE     Molecular cloning and characterization of cDNAs encoding
            ribonucleases in Nicotiana glutinosa leaves that are induced by
            wounding and TMV-infection
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1208)
  AUTHORS   Kimura,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) Makoto Kimura, Kyushu University, Faculty
            of Agriculture; Hakozaki 6-10-1, Higashi-ku, Fukuoka, Fukuoka
            812-8581, Japan (E-mail:mkimura@agr.kyushu-u.ac.jp,
            Tel:81-92-642-2854, Fax:81-92-642-2854)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="Nicotiana glutinosa"
                     /mol_type="mRNA"
                     /db_xref="taxon:35889"
     gene            1. .1208
                     /gene="NGR2"
     mRNA            1. .1208
                     /gene="NGR2"
     CDS             72. .905
                     /gene="NGR2"
                     /codon_start=1
                     /product="RNase NGR2"
                     /protein_id="BAA84468.1"
                     /db_xref="GI:5902454"
                     /translation="MTFLPVHAIVNLRMLFLLIGACCILVGVNGGWDGEVGLLRRGRG
                     GAGGGYQREFDYFKLSLQWPGTYCRRTRRCCSSNACCSRSNSPPVFTIHGLWTEYNDG
                     TWPACCSGKAFDEREISTLLEPLRKYWPSLSCGSPRSCHHRKGPFWAHEWEKHGTCAY
                     PVVHDEYEFFLTTLNIYFKYNVTEVLFEAGYVPSDSEKYPLGGIISSIENAFHATPEL
                     TCSGDALEELRICFYKNFEPRDCARDTSALSRGSCPQYVSLPAHGSWDEIRISNTTAA
                     S"
```

```
BASE COUNT      359 a    232 c    270 g    347 t
ORIGIN

  Query Match           32.0%;  Score 324;  DB 8;  Length 1208;
  Best Local Similarity  67.5%;  Pred. No. 1.2e-71;
  Matches  456;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy         73 GGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTT 132
Db        195 GGACGCGGCGGCGCGAGGAGGAGGATATCAGCGAGAGTTCGATTACTTCAAGCTATCTCTC 254

Qy        133 CAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGC 192
Db        255 CAATGGCCAGGCACCTATTGCCGAAGAACCCGCCGTTGTTGTTCTTCCAACGCCTGTTGC 314

Qy        193 AGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGAT 252
Db        315 AGCCGCTCGAACTCGCCACCTGTATTCACAATTCATGGACTTTGGACAGAGTACAATGAT 374

Qy        253 GGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTG 312
Db        375 GGAACTTGGCCTGCCTGCTGTTCCGGTAAAGCATTTGATGAGAGAGAGATTTCAACATTG 434

Qy        313 ATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAAT 372
Db        435 CTCGAGCCCTTGAGGAAGTATTGGCCTTCTTTAAGTTGTGGTTCCCCCAGATCTTGCCAT 494

Qy        373 GGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCT 432
Db        495 CACAGAAAAGGACCATTCTGGGCACATGAGTGGGAAAAACATGGGACATGTGCTTATCCA 554

Qy        433 GTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAAT 492
Db        555 GTTGTCCATGATGAATATGAGTTCTTTTTGACTACATTGAATATTTACTTCAAGTATAAT 614

Qy        493 GTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTA 552
Db        615 GTTACAGAAGTTCTGTTTGAAGCTGGATATGTACCATCAGATTCCGAAAAGTATCCATTA 674

Qy        553 GGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAA 612
Db        675 GGAGGCATCATTTCATCAATTGAAAACGCTTTTCATGCAACCCCAGAGTTAACATGCTCA 734

Qy        613 AGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGT 672
Db        735 GGCGATGCCCTGGAGGAACTTCGTATATGCTTCTATAAGAATTTTGAGCCTCGTGATTGT 794

Qy        673 GTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAA 732
Db        795 GCACGCGATACTAGCGCCTTATCAAGAGGGTCATGTCCTCAATATGTCAGCTTGCCAGCC 854

Qy        733 TACACGCCATTAGATG 748
Db        855 CATGGATCATGGGATG 870


RESULT 3
AMO315591
LOCUS       AMO315591              1045 bp    mRNA    linear   PLN 10-AUG-2001
DEFINITION  Antirrhinum mollissimum mRNA for S-like RNAse 28 (sl28 gene).
ACCESSION   AJ315591
VERSION     AJ315591.1  GI:15149818
KEYWORDS    ribonuclease; RNase.
SOURCE      Antirrhinum mollissimum
  ORGANISM  Antirrhinum mollissimum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; lamiids; Lamiales; Antirrhinaceae; Antirrhineae;
            Antirrhinum.
REFERENCE   1
  AUTHORS   Liang,L. and Xue,Y.
  TITLE     Molecular characterization of two S RNase-like genes from
            Antirrhinum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1045)
  AUTHORS   Xue,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2001) Xue Y., Plant Genetics and Developmental
            Biology, Institute of Developmental Biology, CAS, 3 Nanyitiao,
            Zhongguancun, Beijing 100080, CHINA
FEATURES             Location/Qualifiers
     source          1. .1045
                     /organism="Antirrhinum mollissimum"
                     /mol_type="mRNA"
                     /cultivar="552"
                     /db_xref="taxon:162390"
                     /clone="pSL28"
                     /tissue_type="pistil"
                     /dev_stage="open flower"
     CDS             33. .863
                     /function="ribonuclease"
                     /codon_start=1
                     /product="S-like RNAse 28"
                     /protein_id="CAC50874.1"
                     /db_xref="GI:15149819"
                     /translation="MAVLTARPLNPAAIQCACFVILWIGLLCVNVGINGSGDLGEKLG
                     ANQRDFDYFHLALQWPGTFCRRTRHCCPTNGCCRGSNAPAEFTIHGLWPDYNDGSWPS
                     CCTGKKFEEKEISTLLGDLNKYWPSLSCGSPSNCHGGKGLFWEHEWEKHGTCSSSVTG
                     AEYNYFVTALKVYFKYNVTEVLREAGYVASNSEKYPLGGIVTAIQNAFHATPELKCSG
                     DAVEELYLCFYKNFEPRDCATKSNKKSCPRYVSLPEYSSLKMANDGNEVSESALDSDI
                     "
BASE COUNT      308 a    204 c    229 g    304 t
ORIGIN

  Query Match           29.1%;  Score 294.4;  DB 8;  Length 1045;
  Best Local Similarity  67.4%;  Pred. No. 4.2e-64;
  Matches  434;  Conservative    0;  Mismatches  201;  Indels    9;  Gaps    1;

Qy        100 CAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGA 159
Db        171 CAGAGAGACTTTGATTACTTCCATCTGGCTCTCCAATGGCCAGGCACATTCTGCCGCAGA 230

Qy        160 ACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTC 219
Db        231 ACCCGCCATTGTTGTCCCACCAATGGCTGTTGTCGAGGTTCAAACGCACCGGCAGAATTT 290

Qy        220 ACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGA 279
Db        291 ACAATCCATGGACTTTGGCCTGATTACAATGATGGAAGTTGGCCTTCATGTTGCACTGGT 350

Qy        280 TCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCT 339
Db        351 AAAAAGTTTGAGGAAAAGGAGATTTCTACATTGCTTGGTGACCTGAATAAGTATTGGCCA 410

Qy        340 AGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCAC 399
Db        411 TCATTAAGCTGTGGTTCTCCGTCAAACTGCCATGGTGGAAAAGGATTATTTTGGGAGCAT 470

Qy        400 GAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTC 459
Db        471 GAGTGGGAAAAACATGGAACTTGTTCGTCTTCAGTCACAGGAGCTGAATACAATTACTTC 530

Qy        460 CTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGC 519
Db        531 GTGACAGCACTTAAAGTTTATTTCAAATACAATGTCACGGAAGTTCTGAGAGAAGCTGGT 590

Qy        520 TATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAAT 579
Db        591 TATGTAGCATCAAATTCTGAAAAATATCCTTTAGGAGGCATCGTTACAGCAATACAAAAT 650

Qy        580 GCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATA 639
Db        651 GCTTTCCATGCGACTCCAGAACTTAAATGCTCGGGTGACGCTGTGGAGGAACTTTACTTG 710

Qy        640 TGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGA 699
```

```
Db        711 TGCTTCTACAAGAATTTTGAGCCTCGTGATTGTG---------CAACCAAATCCAACAAA 761

Qy        700 AAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATT 743
              || || || ||||  || || || ||||| |||||  |  ||||
Db        762 AAATCGTGTCCCAGATATGTCAGCTTGCCAGAATATTCCTCATT 805
```

RESULT 4
AC003000
LOCUS AC003000 92624 bp DNA linear PLN 11-MAR-2002
DEFINITION Arabidopsis thaliana chromosome 2 clone T5I7 map CIC10A06, complete sequence.
ACCESSION AC003000
VERSION AC003000.3 GI:20196974
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 92624)
AUTHORS Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C., Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R. and Venter,J.C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 92624)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 92624)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598383.
FEATURES Location/Qualifiers

```
     source          1. .92624
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="CIC10A06"
                     /clone="T5I7"
     misc_feature    1. .6205
                     /note="overlap with BAC clone F17A14
                     (AC003674:32937. .39141)."
     repeat_region   complement(129. .169)
                     /rpt_family="(CAAAAA)n"
     repeat_region   complement(1000. .1024)
                     /rpt_family="(A)n"
     repeat_region   complement(1266. .1291)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(2779. .2799)
                     /rpt_family="(A)n"
     repeat_region   complement(2880. .2934)
                     /rpt_family="AT_rich"
     repeat_region   3266. .3290
                     /rpt_family="AT_rich"
     gene            complement(4553. .5912)
                     /gene="At2g39710"
                     /note="synonym: T5I7.1; similar to GP 2245012 gnl PID
                     e327000 Z97341; supported by cDNA: gi_15293230_gb_AY051049
                     1_"
     mRNA            complement(<4553. .>5912)
                     /gene="At2g39710"
     CDS             complement(4584. .5912)
                     /gene="At2g39710"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAB87120.2"
                     /db_xref="GI:20196976"
                     /translation="MMDSSSSSSLSLSKNFLRISVLLLIFPLTFCKTSSTNQTLLFSL
                     KTQKLPQSSSDKLSFRHNVTLTVTLAVGDPPQNISMVLDTGSELSWLHCKKSPNLGSV
                     FNPVSSSTYSPVPCSSPICRTRTRDLPIPASCDPKTHLCHVAISYADATSIEGNLAHE
                     TFVIGSVTRPGTLFGCMDSGLSSNSEEDAKSTGLMGMNRGSLSFVNQLGFSKFSYCIS
                     GSDSSGFLLLGDASYSWLGPIQYTPLVLQSTPLPYFDRVAYTVQLEGIRVGSKILSLP
                     KSVFVPDHTGAGQTMVDSGTQFTFLMGPVYTALKNEFITQTKSVLRLVDDPDFVFQGT
                     MDLCYKVGSTTRPNFSGLPMVSLMFRGAEMSVSGQKLLYRVNGAGSEGKEEVYCFTFG
                     NSDLLGIEAFVIGHHHQQNVWMEFDLAKSRVGFAGNVRCDLASQRLGLRD"
     repeat_region   complement(8942. .8968)
                     /rpt_family="AT_rich"
     gene            complement(9879. .11347)
                     /gene="At2g39720"
                     /note="synonym: T5I7.2; supported by full length cDNA:
                     Ceres:253849"
     mRNA            complement(<9879. .>11347)
                     /gene="At2g39720"
     CDS             complement(10039. .11244)
                     /gene="At2g39720"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAB87121.1"
                     /db_xref="GI:2642154"
                     /translation="MASGSYWCYSCSRFVWVSDSISCPDCDGGFLELIQEPLDFTPSD
                     SFTTTTTTQHRSPTRFPPPSSSSSTPSASMHADNSPTPTIVTRTRSNRSPNPVIVLRG
                     SAAAPSSDVVSEGLDRSAFQMYYDDGTDSGLRPLPPSMTEFLLGSGFDRLLDQISQIE
                     LNTNRNLRSCEHPPASKSAIEALPLIEIDPTHLLSDSQSHCAVCKENFVLKSSAREMP
                     CNHIYHPDCILPWLAIRNSCPVCRHELPAEDLTDGTGAALTAVTATAEEEEDSAAGLT
                     IWRLPGGGFAVGRIPGGWRGGDRMMPVVYTEVDGGRLGDERLPRRVAWGSRRGGRDGG
                     GSRERGGGFAGRIMRLFGCFSGSSGSIAAAAAASSGSGSRIRVTRRTRSFSMFSTASS
                     SSRRRNWLA"
     gene            12014. .13082
                     /gene="At2g39725"
                     /note="synonym: T5I7.21; supported by full length cDNA:
                     Ceres: 5605"
     mRNA            join(12014. .12116,12596. .13082)
                     /gene="At2g39725"
     CDS             12603. .12875
                     /gene="At2g39725"
                     /codon_start=1
                     /product="Expressed protein"
                     /protein_id="AAM14851.1"
                     /db_xref="GI:20196975"
                     /translation="MGASKLSGMQKQVLSLYRGFLRAARSRPIEDRKRIEMIVSTEFR
                     HNSKEVDRKNFQYIEYLLRLGTKQLDQLKSPAIVSLSSIKVVTSKT"
     gene            complement(13325. .16081)
                     /gene="At2g39730"
                     /note="synonym: T5I7.18; supported by full length cDNA:
                     Ceres:7114"
     mRNA            complement(join(<13325. .13599,13724. .14005,14087. .14559,
                     14651. .14736,14836. .14941,15033. .15349,15834. .>16081))
                     /gene="At2g39730"
     CDS             complement(join(13484. .13599,13724. .14005,14087. .14559,
                     14651. .14736,14836. .14941,15033. .15349,15834. .15878))
                     /gene="At2g39730"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAB87122.1"
                     /db_xref="GI:2642155"
                     /translation="MAAAVSTVGAINRAPLSLNGSGSGAVSAPASTFLGKKVVTVSRF
                     AQSNKKSNGSFKVLAVKEDKQTDGDRWRGLAYDTSDDQQDITRGKGMVDSVFQAPMGT
                     GTHHAVLSSYEYVSQGLRQYNLDNMMDGFYIAPAFMDKLVVHITKNFLTLPNIKVPLI
                     LGIWGGKGQGKSFQCELVMAKMGINPIMMSAGELESGNAGEPAKLIRQRYREAADLIK
                     KGKMCCLFINDLDAGAGRMGGTTQYTVNNQMVNATLMNIADNPTNVQLPGMYNKEENA
                     RVPIICTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDKIKDEDIVTLV
                     DQFPGQSIDFFGALRARVYDDEVRKFVESLGVEKIGKRLVNSREGPPVFEQPEMTYEK
                     LMEYGNMLVMEQENVKRVQLAETYLSQAALGDANADAIGRGTFYGKGAQQVNLPVPEG
                     CTDPVAENFDPTARSDDGTCVYNF"
     repeat_region   17291. .17350
                     /rpt_family="AT_rich"
     repeat_region   18079. .18106
                     /rpt_family="AT_rich"
     gene            18783. .21237
                     /gene="At2g39740"
```

```
                     /note="synonym: T5I7.4; predicted by genscan"
     mRNA            join(<18783. .18921,19188. .19352,19496. .19674,19766. .19861,
                     19965. .20034,20131. .20261,20593. .>21237)
                     /gene="At2g39740"
     CDS             join(18783. .18921,19188. .19352,19496. .19674,19766. .19861,
                     19965. .20034,20131. .20261,20593. .21237)
                     /gene="At2g39740"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAB87123.1"
                     /db_xref="GI:2642156"
                     /translation="MSRNPFLDPTLQEILQVIKPTRADRDTRITVIDQLRDVLQSVEC
                     LRGATVQPFGSFVSNLFTRWGDLDISVDLFSGSSILFTGKKQKQTLLGHLLRALRASG
                     LWYKLQFVIHARVPILKVVSGHQRISCDISIDNLDGLLKSRFLFWISEIDGRFRDLVL
                     LVKEWAKAHNINDSKTGTFNSYSLSLLVIFHFQTCVPAILPPLRVIYPKSAVDDLTGV
                     RKTAEESIAQVTAANIARFKSERAKSVNRSSLSELLVSFFAKVEDPFEQPVNAARSVS
                     RRNLDRIAQVFQITSRRLVSECNRNSIIGILTGQHIQESLYRTISLPSQHHANGMHNV
                     RNLHGQARPQNQQMQQNWSQSYNTPNPPHWPPLTQSRPQQNWTQNNPRNLQGQPPVQG
                     QTWPVITQTQTQQKSPYKSGNRPLKNTSAGSSQNQGHIGKPSGHMNGVNSARPAYTNG
                     VNSARPPSKIPSQGGQIWRPRHEQ"
     gene            complement(21381. .24860)
                     /gene="At2g39750"
                     /note="synonym: T5I7.5; supported by cDNA:
                     gi_15450748_gb_AY053416 1_"
     mRNA            complement(join(<21381. .21726,21839. .21916,22124. .22246,
                     22328. .22585,22834. .22993,23212. .23322,23481. .23751,
                     23847. .23996,24089. .>24860))
                     /gene="At2g39750"
     CDS             complement(join(21519. .21726,21839. .21916,22124. .22246,
                     22328. .22585,22834. .22993,23212. .23322,23481. .23751,
                     23847. .23996,24089. .24814))
                     /gene="At2g39750"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAB87124.1"
                     /db_xref="GI:2642157"
                     /translation="MKPLTNGDLFKSPTLIKISALVFVTVAFFYLGKHWSDDGYQQLV
                     FFSSSTSGSSIPEVSVSPNSNRVFNLSAIIPTNHTQIEIPATIRQQPPSVVADTEKVK
                     VEANPPPPPPPSPSPPPPPGPVKSFGIVDANGVMSDDFEVGEVESDTVEDWGNQTEIV
                     EAKSDGDSKARVRIKKFGMCPESMREYIPCLDNTDVIKKLKSTERGERFERHCPEKGK
                     GLNCLVPPPKGYRQPIPWPKSRDEVWFSNVPHTRLVEDKGGQNWISRDKNKFKFPGGG
                     TQFIHGADQYLDQMSKMVSDITFGKHIRVAMDVGCGVASFGAYLLSRDVMTMSVAPKD
                     VHENQIQFALERGVPAMAAAFATRRLLYPSQAFDLIHCSRCRINWTRDDGILLLEINR
                     MLRAGGYFAWAAQPVYKHEPALEEQWTEMLNLTISLCWKLVKKEGYVAIWQKPFNNDC
                     YLSREAGTKPPLCDESDDPDNVWYTNLKPCISRIPEKGYGGNVPLWPARLHTPPDRLQ
                     TIKFDSYIARKELFKAESKYWNEIIGGYVRALKWKKMKLRNVLDMRAGFGGFAAALND
                     HKLDCWVLSVVPVSGPNTLPVIYDRGLLGVMHDWCEPFDTYPRTYDFLHASGLFSIER
                     KRCEMSTILLEMDRILRPGGRAYIRDSIDVMDEIQEITKAMGWHTSLRDTSEGPHASY
                     RILTCEKRLLRA"
     repeat_region   24452. .24497
                     /rpt_family="(GGA)n"
     repeat_region   25043. .25065
                     /rpt_family="(A)n"
     repeat_region   25149. .25176
                     /rpt_family="AT_rich"
     gene            25623. .28641
                     /gene="At2g39760"
                     /note="synonym: T5I7.6; predicted by genscan; similar to
                     SP P34568 YNV5_CAEEL;supported by full length cDNA:
                     Ceres:8256"
     mRNA            join(25623. .26134,26620. .26968,27055. .27331,28305. .28641)
                     /gene="At2g39760"
     CDS             join(25746. .26134,26620. .26968,27055. .27331,28305. .28516)

  Query Match             25.8%;  Score 261.2;  DB 8;  Length 92624;
  Best Local Similarity   98.9%;  Pred. No. 1.1e-55;
  Matches 263;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        744 AGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAA 803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35997 AGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAA 36056

Qy        804 AGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTT 863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36057 AGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTT 36116

Qy        864 TACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTAC 923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36117 TACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTAC 36176

Qy        924 TGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCA 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      36177 TGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCA 36236

Qy        984 ACATTTGAGTTCTATGGTAAAAAAAA 1009
              ||||||||||||||||||||||||||
Db      36237 ACATTTGAGTTCTATGGTAAATTGAA 36262


RESULT 5
ATH275979
LOCUS       ATH275979               9423 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Arabidopsis thaliana gene for GDP-mannose pyrophosphorylase, gene
            for vacuolar ribonuclease and gene for proline byosinthesis
            codifying enzyme.
ACCESSION   AJ275979
VERSION     AJ275979.1  GI:13509286
KEYWORDS    GDP-mannose pyrophosphorylase; proline byosinthesis codifying
            enzyme; vacuolar ribonuclease.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Avila,C.
  TITLE     Genes responding to phosphate starvation placed together in
            Arabidopsis genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9423)
  AUTHORS   Avila,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Avila C., Biologia Molecular y Bioquimica,
            Facultad de Ciencias, E-29071, Malaga, SPAIN
FEATURES             Location/Qualifiers
     source          1. .9423
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
     mRNA            join(1405. .1515,1597. .1659,1748. .1856,1883. .2013,
                     2099. .2770)
     CDS             join(1405. .1515,1597. .1659,1748. .1856,1883. .2013,
                     2099. .2770)
                     /codon_start=1
                     /product="GDP-mannose pyrophosphorylase"
                     /protein_id="CAC35355.1"
                     /db_xref="GI:13509287"
                     /translation="MKALILVGGFGTRLRPLTLSFPKPLVDFANKPMILHQIEALKAV
                     GVDEVVLAINYQPEVMLNFLKDFETKLEIKITCSQETEPLGTAGPLALARDKLLDGSG
                     EPFFVLNSDVISEYPLKEILEFHKSHGGEASIMVTKVDEPSKYGVVVMEESTGRVEKF
                     VEKPKLYVGNKINAGIYLLNPSVLDKIELRPTSIEKETFPKIAAAQGHYAMVLPGFWM
                     DIGQPRDYITGLRLYLDSLRKKSPAKLTSGPHIVGNVLVDETATIGEGCLIGPDVAIG
                     PGCIVESGVRLSRCTVMRGVRIKKHACISSSIIGWHSTVGQWARIENMTILGEDVHVS
                     DEIYSNGGVVLPHKEIKSNILKPEIVM"
     exon            1405. .1515
                     /number=1
     intron          1516. .1596
                     /number=1
     exon            1597. .1659
                     /number=2
     intron          1660. .1747
                     /number=2
     exon            1748. .1856
                     /number=3
     intron          1857. .1882
                     /number=3
```

```
    exon            1883..2013
                    /number=4
    intron          2014..2098
                    /number=4
    exon            2099..2770
                    /number=5
    CDS             3288..3473
                    /codon_start=1
                    /product="vacuolar ribonuclease"
                    /protein_id="CAC35356.1"
                    /db_xref="GI:13509288"
                    /translation="MASRLCLLLLVACIAGAFAGDVIELNRSQREFDYFALSLQWPGT
                    YCRGTRHCCSKNACCRG"
    CDS             complement(8081..>8158)
                    /codon_start=1
                    /product="proline biosynthesis enzyme"
                    /protein_id="CAC35357.1"
                    /db_xref="GI:13509289"
                    /translation="MRGKGQVVDGDNGIVYTHQDIPIQA"
BASE COUNT     2795 a   1693 c   1780 g   3155 t
ORIGIN

  Query Match             24.6%;  Score 249;  DB 8;  Length 9423;
  Best Local Similarity   97.4%;  Pred. No. 1.5e-52;
  Matches  264;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy        741 ATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGG 800
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5493 ATCAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGG 5552

Qy        801 AAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAA 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5553 AAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAA 5612

Qy        861 CTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGG--ATATGTATCATC 918
              ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db       5613 CTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATATGTATCATC 5672

Qy        919 ATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGA 978
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       5673 ATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAAAAGATCAGAAAATCCATAAGAAGA 5732

Qy        979 TATCAACATTTGAGTTCTATGGTAAAAAAAA 1009
              ||||||||||||||||||||||||||   ||
Db       5733 TATCAACATTTGAGTTCTATGGTAAATTGAA 5763


RESULT 6
AF139660
LOCUS       AF139660                 981 bp    mRNA    linear   PLN 23-MAR-2000
DEFINITION  Calystegia sepium RNase-like protein mRNA, complete cds.
ACCESSION   AF139660
VERSION     AF139660.1  GI:7288207
KEYWORDS    .
SOURCE      Calystegia sepium (hedge bindweed)
  ORGANISM  Calystegia sepium
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; lamiids; Solanales; Convolvulaceae; Calystegia.
REFERENCE   1  (bases 1 to 981)
  AUTHORS   Van Damme,E.J., Hao,Q., Barre,A., Rouge,P., Van Leuven,F. and
            Peumans,W.J.
  TITLE     Major protein of resting rhizomes of Calystegia sepium (hedge
            bindweed) closely resembles plant RNases but has no enzymatic
            activity
  JOURNAL   Plant Physiol. 122 (2), 433-446 (2000)
  MEDLINE   20144010
   PUBMED   10677436
REFERENCE   2  (bases 1 to 981)
  AUTHORS   Van Damme,E.J.M. and Peumans,W.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1999) Lab. Phytopathology and Plant Protection,
            K.U.Leuven, Willem de Croylaan 42, Leuven, B 3001, Belgium
FEATURES             Location/Qualifiers
     source          1..981
                     /organism="Calystegia sepium"
                     /mol_type="mRNA"
                     /db_xref="taxon:47519"
     CDS             26..787
                     /codon_start=1
                     /product="RNase-like protein"
                     /protein_id="AAF45022.1"
                     /db_xref="GI:7288208"
                     /translation="MASSLPALQIILPLLLLLGTCSANLIRGGHKEFDYFTLALTWSG
                     TECLSVKDSCPTNACSRSEVETGFTIKGLWPDYDDGTWPSCCEGAKYDENEISILSND
                     LSKYWPSYSCMSSSACGSFDASDLAYEWAKHGTCSSPVLGNQYEYFSTALMLYFKYNI
                     SEILSESGYLPSNTAEYKLEGIMSAIQSALRVTPVVKCKSDAVEQVQICFDKTLQLQE
                     CPSTASTCPSLVSLPIKNTIKPLETNNLSFRGMRA"
BASE COUNT      279 a    218 c    195 g    289 t
ORIGIN

  Query Match             18.1%;  Score 182.8;  DB 8;  Length 981;
  Best Local Similarity   57.9%;  Pred. No. 1.1e-35;
  Matches  344;  Conservative    0;  Mismatches  247;  Indels    3;  Gaps    1;

Qy         78 CGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATG 137
              || || |  | ||  | |    || |  ||||| ||||||||| ||||  | |||   ||
Db         91 CGCCAACCTAATCCGTGGCGGACACAAAGAGTTTGATTATTTCACTCTGGCCCTTACTTG 150

Qy        138 GCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGG 197
              | |||| ||  | ||     |   |    | |  ||  |||  || || ||| |    |
Db        151 GTCTGGCACTGAGTGTTTGAGCGTTAAGGACTCCTGTCCCACCAATGCCTGCTCCC---G 207

Qy        198 CTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTC 257
              ||| || | |  ||||  |||||||||  | || || |||||| || ||  | |||||  |
Db        208 CTCTGAAGTTGAAACTGGATTCACAATCAAGGGATTGTGGCCAGATTACGATGATGGGAC 267

Qy        258 GTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGA 317
               ||||| || || ||| |  |  |  | | | | || || ||||| ||||  || || ||
Db        268 TTGGCCGTCTTGCTGTGAGGGTGCCAAATATGACGAAAATGAGATCTCAATTTTTATCGA 327

Qy        318 TGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGG 377
              ||   |    || || |||||          |||   ||  | ||  | ||    |||
Db        328 TGACTTGAGCAAATATTGGCCATCCTATTCCTGTATGTCCTCCTCTGCTTGTGGAAGTTT 387

Qy        378 GAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTT 437
                | |  |||  |||||   | |||||||  ||||| || || ||||| ||||| ||  |
Db        388 CGATGCATCAGATTTGGCATATGAGTGGGCAAAACACGGTACATGTTCCTCTCCCGTACT 447

Qy        438 TCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCAC 497
              |    |  | ||| |  |||||    | ||  | |  || ||||| ||| | ||| |  |
Db        448 TGGAAACCAATATGAATACTTCTCGACCGCTTTGATGCTTTACTTCAAGTACAATATTTC 507

Qy        498 GGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGG 557
              |||  |  | |   || ||| | ||| |  | || |||| ||   | ||    || | |||
Db        508 GGAGATTTTGTCCGAATCTGGATATCTACCATCGAACACTGCTGAATACAAACTGGAAGG 567

Qy        558 TATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGA 617
               ||| |  |||||||||| |  ||  ||    | ||||| |  |||   |||||||| ||
Db        568 CATCATGTCAGCCATTCAAAGCGCTCTTAGGGTAACCCCAGTGGTGAAATGCAAAAGCGA 627

Qy        618 TGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTAAGCCCAGGGACTG 671
              |||  | ||  || | |  ||||||     | |   |  ||  || | | ||
Db        628 TGCTGTTGAGCAAGTTCAGATATGTTTGATAAGACCCTCCAGCTTCAAGAATG 681


RESULT 7
AB112027
LOCUS       AB112027                4210 bp    DNA     linear   PLN 12-JUN-2003
DEFINITION  Nicotiana glutinosa ngr2 gene for ribonuclease NGR2, complete cds.
ACCESSION   AB112027
VERSION     AB112027.1  GI:31620999
KEYWORDS    .
SOURCE      Nicotiana glutinosa
  ORGANISM  Nicotiana glutinosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1
  AUTHORS   Hayashi,T., Kobayashi,D., Kariu,T., Tahara,M., Hada,K., Kozuma,Y.
            and Kimura,M.
  TITLE     Genomic Cloning of Ribonucleases in Nicotiana glutinosa Leaves, as
            Induced in Response to Wounding or to TMV-infection, and
            Characterization of Their Promoters
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4210)
  AUTHORS   Hayashi,T. and Kimura,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2003) Takeshi Hayashi, Kyushu university,
            Laboratory of Biochemistry, Department of Bioscience and
            Biotechnology, Fuculty of Agriculture; Hakozaki 6-10-1, Higashi-ku,
            Fukuoka, Fukuoka 812-8581, Japan
            (E-mail:takeshih@agr.kyushu-u.ac.jp, Tel:81-92-642-4215(ex.4215),
            Fax:81-92-642-2854)
FEATURES             Location/Qualifiers
     source          1. .4210
                     /organism="Nicotiana glutinosa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:35889"
     gene            join(1. .248,1322. .1350,1443. .1516,2308. .2412,3238. .3330,
                     3926. .4210)
                     /gene="ngr2"
     CDS             join(1. .248,1322. .1350,1443. .1516,2308. .2412,3238. .3330,
                     3926. .4210)
                     /gene="ngr2"
                     /codon_start=1
                     /product="ribonuclease NGR2"
                     /protein_id="BAC77612.1"
                     /db_xref="GI:31621000"
                     /translation="MTFLPVHAIVNLRMLFLLIGACCILVGVNGGWDGEVGLLRRGRG
                     GAGGGYQREFDYFKLSLQWPGTYCRRTRRCCSSNACCSRSNSPPVFTIHGLWTEYNDG
                     TWPACCSGKAFDEREISTLLEPLRKYWPSLSCGSPRSCHHRKGPFWAHEWEKHGTCAY
                     PVVHDEYEFFLTTLNIYFKYNVTEVLFEAGYVPSDSEKYPLGGIISSIENAFHATPEL
                     TCSGDALEELRICFYKNFEPRDCARDTSALSRGSCPQYVSLPAHGSWDEIRISNTTAA
                     S"
     intron          2413. .3237
                     /gene="ngr2"
                     /cons_splice=(5'site:NO, 3'site:NO)
BASE COUNT     1175 a    733 c    795 g   1507 t
ORIGIN

  Query Match             10.1%;  Score 102.2;  DB 8;  Length 4210;
  Best Local Similarity   62.9%;  Pred. No. 3.7e-15;
  Matches  158;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        498 GGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGG 557
              ||| || || | | ||||||| |||||  | ||  |  ||||||||||  |||||||
Db       3925 GGAAGTTCTGTTTGAAGCTGGATATGTACCATCAGATTCCGAAAAGTATCCATTAGGAGG 3984

Qy        558 TATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGA 617
              ||| |  || | ||| | || || ||||||   ||||| ||   |    |||  | | ||
Db       3985 CATCATTTCATCAATTGAAAACGCTTTTCATGCAACCCCAGAGTTAACATGCTCAGGCGA 4044

Qy        618 TGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGG 677
              |||  | || ||| | ||||||||||||||||||  ||||| ||||  | || ||||   |
Db       4045 TGCCCTGGAGGAACTTCGTATATGCTTCTATAAGAATTTTGAGCCTCGTGATTGTGCACG 4104

Qy        678 TTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACAC 737
                          ||| |||  |||||| ||  | || || || |||||  |   |
Db       4105 CGATACTAGCGCCTTATCAAGAGGGTCATGTCCTCAATATGTCAGCTTGCCAGCCCATGG 4164

Qy        738 GCCATTAGATG 748
              ⌄    |||  ||||
Db       4165 ATCATGGGATG 4175


RESULT 8
AP006430
LOCUS       AP006430               105384 bp    DNA     linear   PLN 11-JUN-2003
DEFINITION  Lotus japonicus genomic DNA, chromosome 5, clone:LjT33L13, TM0327,
            complete sequence.
ACCESSION   AP006430
VERSION     AP006430.1  GI:31581061
KEYWORDS    HTG.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Asamizu,E., Kato,T., Sato,S., Nakamura,Y., Kaneko,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. IV. Sequence
            Features and Mapping of seventy-three TAC clones which cover the
            7.5 Mb Regions of the Genome
  JOURNAL   DNA Res. (2003) In press
REFERENCE   2  (bases 1 to 105384)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .105384
                     /organism="Lotus japonicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:34305"
                     /chromosome="5"
                     /clone="LjT33L13"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0327"
BASE COUNT    34361 a   18448 c   18695 g   33880 t
ORIGIN

  Query Match             8.2%;  Score 82.6;  DB 8;  Length 105384;
  Best Local Similarity   68.0%;  Pred. No. 3.5e-10;
  Matches  115;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        492 TGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCT 551
              | ||  |  | | ||  || | ||||| |||||| | || ||||  ||||||| ||| ||
Db      58021 TTTCCAGACTCTTCTAAATGAGGCTGGTTATGTTCCATCTAACACAGAAAAGTTTCCCCT 58080

Qy        552 AGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAA 611
              ||||| |  ||| | |||||| |||||||  || |||||  | |||  | |  |||||
Db      58081 TGGAGGCACTGTATCTGCCATTTAGAATGCTTTCCATATGTCTCCTTTAATAATTTGCTC 58140

Qy        612 AAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660
              || || | |  | || ||  | ||  |||||||||||  |  || |||
Db      58141 AAAAGGTTCTGTGGAGGAGCTGCGCCTATGCTTCTATAAGGGCTTCAAG 58189


RESULT 9
AMA489249
LOCUS       AMA489249              4937 bp    DNA     linear   PLN 24-OCT-2002
DEFINITION  Antirrhinum majus X Antirrhinum hispanicum sl28 gene for S-like
            RNase, exons 1-8.
ACCESSION   AJ489249
VERSION     AJ489249.1  GI:21615404
KEYWORDS    S-like RNase; sl28 gene.
SOURCE      Antirrhinum majus x Antirrhinum hispanicum
  ORGANISM  Antirrhinum majus x Antirrhinum hispanicum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; lamiids; Lamiales; Antirrhinaceae; Antirrhineae;
            Antirrhinum.
REFERENCE   1
  AUTHORS   Liang,L., Lai,Z., Ma,W., Zhang,Y. and Xue,Y.
  TITLE     AhSL28, a senescence- and phosphate starvation-induced S-like RNase
            gene in Antirrhinum
  JOURNAL   Biochim. Biophys. Acta 1579 (1), 64-71 (2002)
```

```
  MEDLINE   22288781
   PUBMED   12401221
REFERENCE   2  (bases 1 to 4937)
  AUTHORS   Xue,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2002) Xue Y., Center for Developmental Biology,
            Inst Genet and Develop Biol, CAS, 3 Nanyitiao, Zhongguancun,
            Beijing, 100080, CHINA
FEATURES             Location/Qualifiers
     source          1. .4937
                     /organism="Antirrhinum majus x Antirrhinum hispanicum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:51610"
                     /clone="pGD0428a-e"
                     /haplotype="S2S4"
                     /tissue_type="leaf"
                     /dev_stage="mature plant"
     gene            1. .4915
                     /gene="sl28"
     mRNA            join(1. .236,1289. .1317,1524. .1597,2319. .2424,3141. .3232,
                     4134. .4295,4634. .4709,4860. .4915)
                     /gene="sl28"
     CDS             join(1. .236,1289. .1317,1524. .1597,2319. .2424,3141. .3232,
                     4134. .4295,4634. .4709,4860. .4915)
                     /gene="sl28"
                     /function="RNA metabolism"
                     /codon_start=1
                     /product="S-like RNase"
                     /protein_id="CAD33235.1"
                     /db_xref="GI:21615405"
                     /translation="MALLTARPLNPAAIQCACFVILWIGLLCVNVGINGSGDLGEKLG
                     ANQRDFDYFQLALQWPGTFCRRTRHCCPNNGCCRGSNAPAEFTIHGLWPDYNDGSWPS
                     CCTGKTFEEKEISTLLGDLNKYWPSLSCGSPSNCHGGKGLFWEHEWEKHGTCSSSVTG
                     AEYNYFVTALKVYFKYNVTEVLREAGYVASNSEKYPLGGIVTAIQNAFHATPELKCSG
                     DAVEELYLCFYKNFEPRDCATKSNKKSCPRYVSLPEYSSLKMANDGNEVSESALDSDI
                     "
     exon            1. .236
                     /gene="sl28"
                     /number=1
     intron          237. .1288
                     /gene="sl28"
                     /number=1
     exon            1289. .1317
                     /gene="sl28"
                     /number=2
     intron          1318. .1523
                     /gene="sl28"
                     /number=2
     exon            1524. .1597
                     /gene="sl28"
                     /number=3
     intron          1598. .2318
                     /gene="sl28"
                     /number=3
     exon            2319. .2424
                     /gene="sl28"
                     /number=4
     intron          2425. .3140
                     /gene="sl28"
                     /number=4
     exon            3141. .3232
                     /gene="sl28"
                     /number=5
     intron          3233. .4133
                     /gene="sl28"
                     /number=5
     exon            4134. .4295
                     /gene="sl28"
                     /number=6
     intron          4296. .4633
                     /gene="sl28"
                     /number=6
     exon            4634. .4709
                     /gene="sl28"
                     /number=7
     intron          4710. .4859
                     /gene="sl28"
                     /number=7
     exon            4860. .4915
                     /gene="sl28"
                     /number=8
BASE COUNT     1408 a    841 c    919 g   1766 t      3 others
ORIGIN

  Query Match             7.7%;  Score 78.4;  DB 8;  Length 4937;
  Best Local Similarity   66.7%;  Pred. No. 4.3e-09;
  Matches  112;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        498 GGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGG 557
              ||| || ||     ||||||| ||||| || || ||   |||||| |||||| |||||||
Db       4133 GGAAGTTCTGAGAGAAGCTGGTTATGTAGCATCAAATTCTGAAAAATATCCTTTAGGAGG 4192

Qy        558 TATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGA 617
              ||||| ||||| || || ||||| || |||   || || ||| |   |||     | ||
Db       4193 CATCGTTACAGCAATACAAAATGCTTTCCATGCGACTCCAGAACTTAAATGCTCGGGTGA 4252

Qy        618 TGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAG 665
              ||  | || ||| |     | |||||||| ||  ||||| ||  |||
Db       4253 CGCTGTGGAGGAACTTTACTTGTGCTTCTACAAGAATTTTGAGGTCAG 4300


RESULT 10
AC007195
LOCUS       AC007195               56284 bp    DNA     linear   PLN 27-FEB-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F1P15 map mi398, complete
            sequence.
ACCESSION   AC007195
VERSION     AC007195.6  GI:20198180
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 56284)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 56284)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 56284)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598706.
FEATURES             Location/Qualifiers
     source          1. .56284
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="mi398"
                     /clone="F1P15"
     misc_feature    complement(1. .5693)
                     /note="overlap with BAC clone T24I21 (AC005825:1. .5693)."
     misc_feature    1
                     /note="45452 nt before this point were not included in the
                     submitted sequence, due to overlap with another BAC"
     gene            complement(1221. .2009)
```

```
                /gene="At2g16595"
                /note="synonym: F1P15.2"
mRNA            complement(join(<1221. .1268,1377. .1448,1533. .1736,
                1825. .1902,1989. .>2009))
                /gene="At2g16595"
CDS             complement(join(1221. .1268,1377. .1448,1533. .1736,
                1825. .1902,1989. .2009))
                /gene="At2g16595"
                /codon_start=1
                /product="putative TRAP protein"
                /protein_id="AAM15448.1"
                /db_xref="GI:20198182"
                /translation="MAFAEYRSFSNASVPTSAQATLPYAFAVSKFLQAGAFDFVGTII
                YEIDGKPYQSIFYNGTIEVVEDGPLFRMESVFLSGLLIFVIAFLAIYIQNVLKHMTKK
                TKRAPKVEVGTATKDASHDEWLEGTSYTQSSSKSKKKK"
repeat_region   complement(6149. .6169)
                /rpt_family="AT_rich"
repeat_region   complement(8812. .8849)
                /rpt_family="AT_rich"
repeat_region   8868. .8921
                /rpt_family="AT_rich"
repeat_region   8983. .9012
                /rpt_family="AT_rich"
gene            9751. .10312
                /gene="At2g16590"
                /note="synonym: F1P15.3"
mRNA            9751. .>10312
                /gene="At2g16590"
repeat_region   complement(9759. .10375)
                /rpt_family="Rf:Spacer between 25S and 18S rDNAs part of
                the 10.5 kbp rDNA repeat#Rf:ATR0032|X15550 intergenic
                spacer between 25S and 18S rDNAs"
repeat_region   complement(9765. .10375)
                /rpt_family="Rf:rDNA repeat of 18S rDNA, intergenic
                spacer, 5.8S rDNA, spacer, 25S rDNA, spacer entire repeat
                unit is 10.5 kbp, 2 loci of ~3.5 Mbp on chromosomes II and
                IV, 670 copies#Rf:ATR0026|X52322|X15550 10.5 kbp rDNA
                repeat"
CDS             9992. .10312
                /gene="At2g16590"
                /codon_start=1
                /product="unknown protein"
                /protein_id="AAD26488.1"
                /db_xref="GI:4589971"
                /translation="MEPTNTSYSITTHTPNVQLTQNHQSFQLIYRSTRLRRKKTDESE
                PKPYKREIPSGQPRRIHAFQRSTPIQPRRDKGSHFVHHPWPAFEQPDSHQKMVAKNIF
                VTVC"
repeat_region   complement(10151. .10375)
                /rpt_family="Rf:rDNA repeat of 18S rDNA, intergenic
                spacer, 5.8S rDNA, spacer, 25S rDNA, spacer entire repeat
                unit is 10.5 kbp, 2 loci of ~3.5 Mbp on chromosomes II and
                IV, 670 copies#Rf:ATR0026|X52322|X15550 10.5 kbp rDNA
                repeat"
repeat_region   complement(10151. .10211)
                /rpt_family="Rf:Spacer between 25S and 18S rDNAs part of
                the 10.5 kbp rDNA repeat#Rf:ATR0032|X15550 intergenic
                spacer between 25S and 18S rDNAs"
repeat_region   complement(10151. .10209)
                /rpt_family="Rf:rDNA repeat of 18S rDNA, intergenic
                spacer, 5.8S rDNA, spacer, 25S rDNA, spacer entire repeat
                unit is 10.5 kbp, 2 loci of ~3.5 Mbp on chromosomes II and
                IV, 670 copies#Rf:ATR0026|X52322|X15550 10.5 kbp rDNA
                repeat"
repeat_region   10939. .10958
                /rpt_family="(TAAAA)n"
repeat_region   11493. .11517
                /rpt_family="AT_rich"
repeat_region   11574. .11608
                /rpt_family="AT_rich"
repeat_region   complement(11640. .11693)
                /rpt_family="AT_rich"
repeat_region   complement(12610. .12653)
                /rpt_family="AT_rich"
repeat_region   complement(14318. .14338)
                /rpt_family="(GA)n"
gene            14379. .14705
                /gene="At2g16580"
                /note="synonym: F1P15.4"
mRNA            <14379. .>14705
                /gene="At2g16580"
CDS             14379. .14705
                /gene="At2g16580"
                /codon_start=1
                /product="putative auxin-induced protein"
                /protein_id="AAD26489.1"
                /db_xref="GI:4589972"
                /translation="MSILKKSTKLAQTAMLRQILKRCSSLGKKNGGGGYEEVDLPLDV
                PKGHFPVYVGHNRSRYIVPISFLTNLDFQCLLRRAEEEFGFDHDMGLTIPCDELFFQD
                LTSMIR"
repeat_region   17248. .17293
                /rpt_family="AT_rich"
repeat_region   17712. .17837
                /rpt_family="AT_rich"
gene            19114. .21162
                /gene="At2g16570"
                /note="synonym: F1P15.5; identical to
                GP:469193:gnl:PID:d1006574:D28868; supported by cDNA:
                gi_469192_dbj_D28868 1_ATHAMID"
mRNA            19114. .21162
                /gene="At2g16570"
CDS             19183. .20883
                /gene="At2g16570"
                /codon_start=1
                /product="amidophosphoribosyltransferase"
                /protein_id="AAD26498.1"
                /db_xref="GI:4589981"
                /translation="MAATTSFSSSLSLITKPNNSSYTNQPLPLFPKPFLKPPHLSLLP
                SPLSSPPPSLIHGVSSYFSSPSPSEDNSHTPFDYHNDEDDEKPREECGVVGIYGDPEA
                SRLCYLALHALQHRGQEGAGIVTVSPEKVLQTITGVGLVSEVFNESKLDQLPGEFAIG
                HVRYSTAGASMLKNVQPFVAGYRFGSIGVAHNGNLVNYKTLRAMLEENGSIFNTSSDT
                EVVLHLIAISKARPFFMRIIDACEKLQGAYSMVFVTEDKLVAVRDPYGFRPLVMGRRS
                NGAVVFASETCALDLIEATYEREVYPGEVLVVDKDGVKSQCLMPKFEPKQCIFEHIYF
                SLPNSIVFGRSVYESRHVFGEILATESPVECDVVIAVPDSGVVAALGYAAKSGVPFQQ
                GLIRSHYVGRTFIEPSQKIRDFGVKLKLSPVRGVLEGKRVVVVDDSIVRGTTSSKIVR
                LLREAGAKEVHMRIASPPIVASCYYGVDTPSSEELISNRLSVEEINEFIGSDSLAFLS
                FDTLKKHLGKDSKSFCYACFTGDYPVKPTEVKVKRGGGDFIDDGLVGSFENIEAGWVR
                "
repeat_region   complement(19258. .19385)
                /rpt_family="(GGAGAA)n"
repeat_region   21487. .21535
                /rpt_family="AT_rich"
repeat_region   22325. .22397
                /rpt_family="AT_rich"
repeat_region   22477. .22509
                /rpt_family="AT_rich"
repeat_region   23028. .23090
                /rpt_family="AT_rich"
repeat_region   23123. .23253
                /rpt_family="AT_rich"
repeat_region   23419. .23679
                /rpt_family="AT_rich"
repeat_region   23717. .24265
                /rpt_family="Rf:ATREP1|ATREP1 ATREP1 dispersed repeat from
                GB|D86608 - a consensus."
repeat_region   23808. .23845
                /rpt_family="AT_rich"
repeat_region   23916. .23965
                /rpt_family="AT_rich"
repeat_region   complement(24055. .24082)
                /rpt_family="AT_rich"
repeat_region   24130. .24267
                /rpt_family="AT_rich"
repeat_region   24397. .24441
                /rpt_family="AT_rich"
repeat_region   24485. .24529
                /rpt_family="AT_rich"
```

```
     gene            complement(24618. .28304)
                     /gene="At2g16560"
                     /note="synonym: F1P15.6; pseudogene, reverse
                     transcriptase"
                     /pseudo
     repeat_region   28430. .28467
                     /rpt_family="AT_rich"
     repeat_region   complement(28430. .28467)
                     /rpt_family="AT_rich"
     repeat_region   28602. .28640
                     /rpt_family="AT_rich"

  Query Match              5.5%;  Score 55.2;  DB 8;  Length 56284;
  Best Local Similarity   73.7%;  Pred. No. 0.0034;
  Matches  112;  Conservative    0;  Mismatches   33;  Indels    7;  Gaps    3;

Qy         27 TTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGGAGACGTCATCGA 86
              | | | | || |||||||||||| ||| |||| ||| |  |||   |||||| ||| || |
Db       7060 TGTTTTTTTTTCCTTCTCGTTTCGTTTATCACCGAACGATT--GCGGAGATGTCGTCAA 7117

Qy         87 ACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAAC 146
              | ||||||| |||| |   | ||||||||||||||  ||  |||||||||| | || |||
Db       7118 AGTCAATCGCTCTCCCATAAAATTCGATTATTTCG--CTGGCTCTTCAATGACATGAAAC 7175

Qy        147 CTATTGCCGTGGAACTCGCCATTGTTGCTCCA 178
              ||| |||    |||||| |||||| ||||| |
Db       7176 CTACTGC---TGAACTCACCATTGATGCTCAA 7204


RESULT 11
AP003349/c
LOCUS       AP003349              142955 bp    DNA     linear   PLN 29-MAR-2003
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            PAC clone:P0674H09.
ACCESSION   AP003349 BA000010
VERSION     AP003349.4  GI:22093582
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
            Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,
            Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,
            Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
            Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
            Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T.,
            Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
            Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
            Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
            Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
            Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
            Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
            Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
            Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
            Yano,M., Jiang,J. and Gojobori,T.
  TITLE     The genome sequence and structure of rice chromosome 1
  JOURNAL   Nature 420 (6913), 312-316 (2002)
  MEDLINE   22337376
   PUBMED   12447438
REFERENCE   2  (bases 1 to 142955)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Aug 2, 2002 this sequence version replaced gi:20975423.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SlicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0674H09 clone has an overlap with P0696G06 clone
            (DDBJ : AP003316) at the position 1 to 51,418 of 5' end and an
            overlap with P0506A10 clone (DDBJ: AP003418) at the position 55,654
            to 142,955 of 3' end. Detailed information on overlap and assembly
            quality together with annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES            Location/Qualifiers
     source         1. .142955
                    /organism="Oryza sativa (japonica cultivar-group)"
                    /mol_type="genomic DNA"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:39947"
                    /chromosome="1"
                    /clone="P0674H09"
     gene           5794. .6351
                    /gene="P0674H09.1"
     CDS            5794. .6351
                    /gene="P0674H09.1"
                    /note="hypothetical protein"
                    /codon_start=1
                    /protein_id="BAC06879.1"
                    /db_xref="GI:22093583"
                    /translation="MGKRRPRCGSARHGVARCLAWWRLARRHGGRRWRMGEGKGLAAM
                    EAVGAKEAVADVGDIDCFSLLMPNGLLLLLLARPSRLPSSEFCRKLPIKNGYRPNGRK
                    KPHIISTTISERDYRFHFRYRQLPIPTETVGANRKRKPEFSEISVPFSPLGRTRWCLV
                    RGAGRRLAARLRSRVIYAIWAAESR"
     gene           join(8132. .8265,8497. .9025)
                    /gene="P0674H09.2"
     CDS            join(8132. .8265,8497. .9025)
                    /gene="P0674H09.2"
                    /note="hypothetical protein"
                    /codon_start=1
                    /protein_id="BAC06880.1"
                    /db_xref="GI:22093584"
                    /translation="MADSARRDRSGLGPRKRKERVRPESTSRVLARAGSSDTRNDKRR
                    RAVAQQEGEASVRERKAGNGGGEREVEPCRFGEGEGWEREALPPDFGRTRAAQLGHAR
                    WLAGPSGSATTGAARQRGAGAARAATARARARGTVCGDAKFAGGGSPGRERWGEGERE
                    NEREEKALGLAWRRLTRTRVQRMGRARGKLRERERERREMGREINRPTEPEGGKLDFYR
                    GI"
     gene           join(14423. .14497,14878. .15690,15782. .15901,16516. .17199)
                    /gene="P0674H09.3"
     CDS            join(14423. .14497,14878. .15690,15782. .15901,16516. .17199)
                    /gene="P0674H09.3"
                    /note="contains EST D39962(S1657)
                    similar to Arabidopsis thaliana chromosome 3, At3g59690
                    unknown protein"
                    /codon_start=1
                    /protein_id="BAC06881.1"
                    /db_xref="GI:22093585"
                    /translation="MGKKGGWITALKKAFTSGPKDKPTNGQLVAQYSHQHRSGGGGAA
                    AARDKKRWGFGRSRQHAEPSPAGALINIPLYREPSSIEKILVDAEMEQQHRQYRAHYQ
                    ITPAKPATVTAPAAAASLPAPPPPVSARERQRRRDVDDDKPAIVLPLPPPSPPPLIRR
                    FDHDREQQQKLQQMQQQTRAETEWRRPQAQRRRAARPRGPAPPDRARAAAVAIQSAFR
                    GYMARRNYRSLRGLIRLQGVVRGPSVRRQTAHAMRCMQMLVRVQSQVRASRVEAMERR
                    NRHHHAAMLRDAARWRAASQDGGIWEDSLLSRDEMDARTKRKVEAVIKRERALAYAYS
                    HQLLKATPMAAHAILADLQSGRNPWWWTPIERRHEAEAAAALLSRQRVEHVGNGGGGG
                    RHVAVTATPARSVVSSYSTATATTAARPRATRPAKVASSYGGGGSIRDDESLTSCPAF
                    GGALPNYMTPTPMSASAKARARAQMLRQQQEKQAQAAQEKPRFSFGLGQSIGSWAKSPF
```

WKGGAAAAPPSISSRVATPAASVAGRHRPTRSSVSELSVDSAVSMPAGIGRRTFK"
gene join(20555. .21175,21273. .21372,21678. .21822,22501. .22624,
23076. .23609)
/gene="P0674H09.4"
CDS join(20555. .21175,21273. .21372,21678. .21822,22501. .22624,
23076. .23609)
/gene="P0674H09.4"
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 1, F5I14.2"
/codon_start=1
/protein_id="BAC06882.1"
/db_xref="GI:22093586"
/translation="MEGGKLLGVAHPEPANNIDADLRYDLGQSRMQVDGPVVLNRSAE
LEPSDSMAIDDVPVEASSQPAPAKQSPALMDTIVEVQKQLKRKRASSGPALAAADKDA
LVAGCCQELEGLLEYYREVSGHRMQFEVGNLSTNAAIGCLLEESSLGLSKLVDEIYEK
LKGMEGVSATSVRSSVLLIGQRMMYGQSSPDADVLEDESETALWCWEVRDLKVIPLRM
RGPLSTRRTARKKIHERITAIYSTLSVLEAPGAEAQVNDMRKASLKLSKALNLEGIKS
LVERATQKSNIERGAKNTGSTAKEPMQEMVKSNNDTGIIENVDDSQLQKNTSTNEKDT
QKAQKQVEKELKQKEKEEARMRKQQKKQQEEALREQKRREKEEAEMKKQQRKQEEEAQ
KEQKRREKEEAETRKQQKKQQEEAEKEQKRREKEAVQLKKQLAIQKQASMMERFFKNK
KDSEKLEKPGGKDSGVQTTDPCTTNKEVVPLVTSIIDSSFSQKENWALEDLRRLISF"
gene join(23955. .24091,24190. .24271,24342. .24477,24697. .24737,
24942. .25340,25423. .25560,25696. .25741,25833. .26002)
/gene="P0674H09.5"
CDS join(23955. .24091,24190. .24271,24342. .24477,24697. .24737,
24942. .25340,25423. .25560,25696. .25741,25833. .26002)
/gene="P0674H09.5"
/codon_start=1
/product="putative FAS1"
/protein_id="BAC06883.1"
/db_xref="GI:22093587"
/translation="MLPAVELQFHGTNHANPLPTRSIKRKLLQFDKSNRPAYYGTWRK
KSAVVGPRCPLKMDPDLDYEVDSDDEWEEEDPGESLSDCEKDNDEVMEEDSKITDEES
EDSFFVPDGYLSDNEAKNKIVEQDKSLYEYGIQIESLLDDKDEASSSPPDQCAEVEEF
RALLRQQKVLNTLTEQALRKSQPLVISNLTHEKAELLTAGDLKGTSKIEQLCLQVLSM
RICPGGATIDLPVIDSSSANAEETNQLNVKSSPAAASAIPDTDLAEIVKVIGSCRDGI
NKLVESLHQKFPNVSKSQLKNKVREISEFVDNRWQVKKEVLSKLGLSSSPASSKKPKS
IATYFSKRCLPPEEAILASPELRLKSKTTQNVNGDTDIPRINLLPSSQ"
gene complement(26516. .27016)
/gene="P0674H09.6"
CDS complement(26516. .27016)
/gene="P0674H09.6"
/note="contains EST C27797(C52876)
unknown protein"
/codon_start=1
/protein_id="BAC06884.1"
/db_xref="GI:22093588"
/translation="MAAGKGSHTHRAFLLCNYALLGAASSCIFLTLSLRLLPSPCGLL
LLFLHALTAVFSAAGCSGSFTAPATPAQWHNAHTAGAALTAIFQGAVALLAFTRTSDF
LAELQSYVRDDDAAVILKMVGGLGTAIFVLEWAALALAFSLRLDDDDGAGDYDNKNWA
AASYHV"
gene complement(join(28586. .28757,28943. .29001,29174. .29220,
29897. .29927,30140. .30202,30393. .30488,30572. .30655,
30745. .30837,31138. .31215,31571. .31661,31741. .31841,
31936. .32754,32849. .32899))
/gene="P0674H09.7"
CDS complement(join(28586. .28757,28943. .29001,29174. .29220,
29897. .29927,30140. .30202,30393. .30488,30572. .30655,
30745. .30837,31138. .31215,31571. .31661,31741. .31841,
31936. .32754,32849. .32899))
/gene="P0674H09.7"
/note="contains ESTs AU057282(S21302),AU057281(S21302)
similar to Arabidopsis thaliana chromosome 3, At3g59780
unknown protein"
/codon_start=1
/protein_id="BAC06885.1"
/db_xref="GI:22093589"
/translation="MLLPVCSAAPTCSPLCPVTSSVAHVVGAPALRGRVVEVSALRSY
ADPLVARPVPDRPLLADSSILSPYSASPDDIVRGLAAAEVPSADATAAGDALCCGGVE
PATLTAVSDLAATTADAAVTDAAERALLDAPVPTTFPADASDAEAAFARLIDSLGKKI
FQAEDALTEGYDKLRLSAYQSLSAWRKSVDGAVGGLTASVDATKKQAAGGVTDASGAL
QDKVAGAGTVAVDVLRKAIVAAEDSLGSAATSIGYYYGSTKSSLPPNVKDLLNSSEEK
ASLVLKPIGGALQQVYVVVEGIEKNVGLDPSDPIVQLAVLLGGSTTIGLTYWLFAYGG
YSGDLSPESTLELLRNDDKADLRVKDGIPDLRRAARSKYATAASPEIKGSVKRLLKSG
RDVDDALLAVIIRNLKLVKGDSKVIIMDANGSRSKAIARLLKKLGVQRPYLVKGGFQS
WAKNLRVKELKPETALTVINEDAEEILEDIKPTPTLVLGSLLGLSAAAYALLEWETTL
QYIGVLSLGLTIYLRFSTYEGSEDLEQDLKVGAEAFSWAAKKLEPSKIGLATSPSTTA
VQDRVLQAAAKHESQPSDAEADSLVSEA"
gene complement(join(34221. .34296,34581. .34647,35201. .35291))
/gene="P0674H09.8"
CDS complement(join(34221. .34296,34581. .34647,35201. .35291))
/gene="P0674H09.8"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC06886.1"

Query Match 5.4%; Score 54.6; DB 8; Length 142955;
Best Local Similarity 53.5%; Pred. No. 0.0047;
Matches 114; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy      269 GTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTGAGA 328
            ||   |||  |||||  ||||    |    |||| || | ||||| |       || |||
Db    63672 GTGTGTATTTATCTGGTTTTACCTTGTGACAGATCTCGAGGTTGAAGACGATACTCGAGG 63613

Qy      329 AGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGTCAT 388
            ||||||||||     |    || || |||   ||| | |||  |||||||||| || |||
Db    63612 AGTACTGGCCGTCCTTGTACTGCGGCTCTTTCTCAACCTGCTTTGGTGGGAAAAGGCCAT 63553

Qy      389 TTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGT 448
            ||||||  || |||   || |    |              | |    |||      | ||
Db    63552 TTTGGGTGCATGAGGTAGACATTGGTATCTAAATGAGATATGTGATTTGATGAGGAGAAA 63493

Qy      449 ATAATTACTTCCTTACCACACTTAATCTCTACT 481
              || | || ||  |  ||    ||  |||  |||
Db    63492 TTACTGACATCTTACCCTTTTTTCCTCTTGACT 63460
```

RESULT 12
AP003418/c
LOCUS AP003418 145265 bp DNA linear PLN 29-MAR-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0506A10.
ACCESSION AP003418 BA000010
VERSION AP003418.2 GI:15216346
KEYWORDS .
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T., Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y., Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K., Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K., Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y., Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Itoh,S., Itoh,T., Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S., Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T., Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M., Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M., Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T., Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K., Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K., Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y., Yano,M., Jiang,J. and Gojobori,T.
TITLE The genome sequence and structure of rice chromosome 1
JOURNAL Nature 420 (6913), 312-316 (2002)
MEDLINE 22337376
PUBMED 12447438
REFERENCE 2 (bases 1 to 145265)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

```
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Aug 20, 2001 this sequence version replaced gi:13366121.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
             Detailed information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES            Location/Qualifiers
     source         1. .145265
                    /organism="Oryza sativa (japonica cultivar-group)"
                    /mol_type="genomic DNA"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:39947"
                    /chromosome="1"
                    /clone="P0506A10"
     gene           complement(join(2525. .2595,2686. .2824,3342. .3503,
                    4105. .4206,4353. .4426,4567. .4595,4746. .4969))
                    /gene="P0506A10.1"
     CDS            complement(join(2525. .2595,2686. .2824,3342. .3503,
                    4105. .4206,4353. .4426,4567. .4595,4746. .4969))
                    /gene="P0506A10.1"
                    /note="contains ESTs C96756(C10648),D22272(C10648)"
                    /codon_start=1
                    /product="putative S-like ribonuclease RNS2"
                    /protein_id="BAB86068.1"
                    /db_xref="GI:19386686"
                    /translation="MATARKAPRMVAIWAVVVVVVVAADLFGAAVARSASARHAVGKKQ
                    REFDYFALALQWPGTICASTRHCCAINGCCRSEPLQTFTIHGLWPDYDDGTWPACCRH
                    TSFDMDKILPLKPTLEKYWPSLYCSSSSTCFSGKGPFWAHEEMLASGGIHVSNGKQYA
                    LTDVIDAIKCAFGASPQIVCKKGSVEELRLCFDKDLKKTCTILLNSVVYFQPLDCLTT
                    TATNENVSKKKYCPRYITLPTYDPIVHANSTREIITVESEVYGYLYTS"
     gene           complement(join(5975. .6058,6439. .6510,6901. .7062,
                    7248. .7304,7886. .7987,8137. .8210,8302. .8330,8428. .8648))
                    /gene="P0506A10.2"
     CDS            complement(join(5975. .6058,6439. .6510,6901. .7062,
                    7248. .7304,7886. .7987,8137. .8210,8302. .8330,8428. .8648))
                    /gene="P0506A10.2"
                    /note="contains ESTs AU092918(C60895),C28387(C60895)"
                    /codon_start=1
                    /product="putative S-like ribonuclease RNS2"
                    /protein_id="BAB86069.1"
                    /db_xref="GI:19386687"
                    /translation="MEEKRQRLVISLSALCLAVAVMSGALLPPRASAAGGVGRKRRWA
                    GFDYYVLALQWPGTVCRQTSHCCSSNGCCRSHPLKFFTIHGLWPQYSYGGWPSCCRPT
                    TFDGNKISRLKTILEEYWPSLYCGSFSTCFGGKRPFWVHEDEYDYFSTALYLYSKYNV
                    TKALKKAHIYPRGGRKYLVGHIVSAIEFSFGAMPSIVCKNGSVQELRLCFHKDYQLRV
                    RTLLEETIALDTSLYHHINHTHLAILQKESVTKSTLSTNHISRSHISL"
     gene           join(11526. .11584,11964. .13698)
                    /gene="P0506A10.3"
     CDS            join(11526. .11584,11964. .13698)
                    /gene="P0506A10.3"
                    /note="hypothetical protein
                    similar to Arabidopsis thaliana chromosome 3, At3g02650"
                    /codon_start=1
                    /protein_id="BAB86070.1"
                    /db_xref="GI:19386688"
                    /translation="MFHGNADNSIEDLDSNLSRNYHAAAAAAADEEEMWRRSGARALL
                    LLRSSIHRPLPPPPSPLRALTRAPPPPRPLSRFLSSSPEPLPDAGGGAASASSDPTDA
                    GEDGLASDPTEAGEDNLASLWEEDAGDADDVFAADASPDPADAVVDEVLAARVGAVVE
                    STPEDQIPSALADMVVDFNEPLLAAILVSAAESCSGKKLLRLFKSAGKNNPAAKSLAN
                    LEIVVDKIADSDEIDKMDVYLLWDLVKEMGAVPGSVNTQVLNKLLAMFWKLEKSKAAL
                    EAFDMFSELGCAPDGDSYYLAIQAAGRKSMVDAAWRVCDGMISSGCFPDGEKVGDIVM
                    FFCKKKKVKEAHSVYITAKEKKVQAPTSALTFLVSALARNDETINTALELLEEYQGES
                    LKDAGQSYAAVIHGLCNTNNVEDAKKVLMRMVNLGPAPGSAVFNFVITALSKNGEMED
                    AKGLMRVMENQGICPDIYTYSVIMSGYAKGGMMDEAHALLRDAKRIHPRLSRVTYHTL
                    IRGYCKMEEFEKALECLKEMKEDGMEPNMDEYNKLIQSLCLKALDWRTAENLLKEMED
                    GGLCLKGTTRSLIAAVKELEMDELSKASQEA"
     gene           join(16095. .16181,16234. .16414,17357. .17415,17616. .17691,
                    17836. .17910,17987. .18074,18241. .18466,18669. .18784,
                    19152. .19363,19862. .19896,19985. .20087,20625. .20733,
                    20818. .21016)
                    /gene="P0506A10.4"
     CDS            join(16095. .16181,16234. .16414,17357. .17415,17616. .17691,
                    17836. .17910,17987. .18074,18241. .18466,18669. .18784,
                    19152. .19363,19862. .19896,19985. .20087,20625. .20733,
                    20818. .21016)
                    /gene="P0506A10.4"
                    /note="contains ESTs D41563(S4135),C98908(E3241)"
                    /codon_start=1
                    /product="putative beta-glucosidase"
                    /protein_id="BAB86071.1"
                    /db_xref="GI:19386689"
                    /translation="MWGHLPGRQSLRGGSHRLVPVPPTPVAAHAETTRGARLTCRATR
                    RARCGDRLLLAAAAGGNGEHGARRGGDPRRLPRRLRLRRRHLRLPERILDGSSGEVAV
                    DHYHRYKEDIELMASLGFRAYRFSISWPRIFPDGLGKNVNEQGVAFYNDLINFMIEKG
                    IEPYATLYHWDLPHNLQQTVGGWLSDKIVEYFALYAEACFANFGDRVKHWITINEPLQ
                    TAVNGYGIGHFAPGGCEGETARCYLAAHYQILAHAAAVDVYRRKFKAVQGGEVGLVVD
                    CEWAEPFSEKTEDQVAAERRLDFQLGWYLDPIYFGDYPESMRQRLGDDLPTFSEKDKE
                    FIRNKIDFVGINHYTSRFIAHHQDPEDIYFYRVQQVERIEKWNTGEKIGERAASEWLF
                    IVPWGLRKLLNYAAKRYGNPVIYVTENGMDEEDDQSATLDQVLNDTTRVGYFKGYLAS
                    VAQAIKDGADVRGYFAWSFLDNFEWAMGYTKRFGIVYVDYKNGLSRHPKASARWFSRF
                    LKGDDAENKADMN"
     gene           complement(join(22252. .22360,22835. .22882,23187. .23197))
                    /gene="P0506A10.5"
     CDS            complement(join(22252. .22360,22835. .22882,23187. .23197))
                    /gene="P0506A10.5"
                    /note="hypothetical protein"
                    /codon_start=1
                    /protein_id="BAB86072.1"
                    /db_xref="GI:19386690"
                    /translation="MNDRNFLIGDMSLTGYKDVRSEYIPDRAISLRCSEKGSGLCMNQ
                    SKGSQKQDARE"
     gene           complement(join(25382. .26209,26300. .26547,27139. .27344,
                    27434. .29034))
                    /gene="P0506A10.6"
     CDS            complement(join(25382. .26209,26300. .26547,27139. .27344,
                    27434. .29034))
                    /gene="P0506A10.6"
                    /note="contains EST D24760(R2517)"
                    /codon_start=1
                    /product="putative formin-like protein AHF1"
                    /protein_id="BAB86073.1"
                    /db_xref="GI:19386691"
                    /translation="MPSLRRFLELVVVGIVVCGVNGGSDGLEVAVARRQLHQPFFPDQ
                    SSSPPTPAPPGPAPPFFPALPVPPPPPATAGQEQPTYPALVLPNTGAGGAAATAAPDG
                    GGGGGGGARKSKSSASKLVPAIVLPLLTVAVLGLSIAFFFTHRRGNAARGGGGGGGCV
                    GGGDAKFLHPERASLFARDEFGGSGGAAAPPAAAMDYRYVGNAGIGRMDEKSSETTSS
                    GDEASRSTGGSPELRPLPPLLARQCGPMGARSPGSGVGGFASPSSGDEEFYSPQGSSK
                    MSTSHRTLAAAVEAAVAARDRSKSPSPGSIVSTPSYPSSPGATMSPAPASPPLFSSPG
                    QSGRRSVKSRSDSVRTFGQPPAPPPPPPFAPTLPPPPPPRRKPPSPSPPSSPLIENTS
                    ALRSTTTTDTTIPRNPFVQPPPPPTHTHGPPPPPPPPPPPPVGYWESRVRKPGTGTSK
                    ETRSPALSPPPQAASFKSGLPTDAFPGRLADNADHAAAAAAGGGGDKSEETTPRPKLK
                    PLHWDKVRASSDRVMVWDQLKSSSFQVNEEMIETLFICNPANSAPPKEPATRRPVLPT
                    PKTDNKVLDPKKSQNIAILLRALNVSKEQVCDALCEGNTENFGAELLETLLKMAPTKE
                    EEIKLREFKEETSPIKLGPAEKFLKAVLDIPFAFKRVDAMLYIANFESEVNYLKKSFE
                    TLETACDELRNSRLFLKLLEAVLKTGNRMNVGTNRGDAHAFKLDTLLKLVDVKGTDGK
                    TTLLHFVVQEIIRTEGSHLSASNQSTPRTQANPLRDELECKKLGLQVVAGLGNELSNV
                    KKAAAMDSDVLSSYVSKLAGGIEKITEVLRLNEEVKSREDAWRFHDSMQKFLKRADDD
                    IIRVQAQESVALSLVKEITEYFHGDSAKEEAHPFRIFMVVRDFLSVLDQVCKEVGRIN
                    DRTIASSVRHFPVPVNPMMPQLFPRIHALRAGISDDESSATSASSP"
```

```
     gene            join(29971. .30027,30305. .30501,30556. .30748)
                     /gene="P0506A10.7"
     CDS             join(29971. .30027,30305. .30501,30556. .30748)
                     /gene="P0506A10.7"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB86074.1"
                     /db_xref="GI:19386692"
                     /translation="MHACKRKPKALTVAGSIEFSAITSGTEGIRWQSNVPEPPNVGRV
                     RKKGRSRVRLRVNLTYSINPCPASACVRLPMPKSPDASASRRSDPHVRDSSTFTRGSI
                     WRKRRRRACVARAGDRSIMDACIELPCPCIPGYRAGPEEAGRGAVK"
     gene            join(35765. .35828,35844. .36037,36067. .36285)
                     /gene="P0506A10.8"
     CDS             join(35765. .35828,35844. .36037,36067. .36285)
                     /gene="P0506A10.8"
                     /note="hypothetical protein"
                     /codon_start=1

  Query Match             5.4%;  Score 54.6;  DB 8;  Length 145265;
  Best Local Similarity   53.5%;  Pred. No. 0.0047;
  Matches  114;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        269 GTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTGAGA 328
              ||   |||  |||||  ||||    |    |||| || | ||||| |      || |||
Db       8019 GTGTGTATTTATCTGGTTTTACCTTGTGACAGATCTCGAGGTTGAAGACGATACTCGAGG 7960

Qy        329 AGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGTCAT 388
              ||||||||||      |    || ||| |||  ||| | ||| ||||||||| || |||
Db       7959 AGTACTGGCCGTCCTTGTACTGCGGCTCTTTCTCAACCTGCTTTGGTGGGAAAAGGCCAT 7900

Qy        389 TTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGT 448
              ||||||  || |||   || |    |            | |    |||      | ||
Db       7899 TTTGGGTGCATGAGGTAGACATTGGTATCTAAATGAGATATGTGATTTGATGAGGAGAAA 7840

Qy        449 ATAATTACTTCCTTACCACACTTAATCTCTACT 481
               || | || || |  ||     ||  |||  |||
Db       7839 TTACTGACATCTTACCCTTTTTTCCTCTTGACT 7807


RESULT 13
PFMAL6P1_13
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL6P1 Accession AL844505
     Fragment Name      Begin     End
     PFMAL6P1_00            1    110000
     PFMAL6P1_01       100001    210000
     PFMAL6P1_02       200001    310000
     PFMAL6P1_03       300001    410000
     PFMAL6P1_04       400001    510000
     PFMAL6P1_05       500001    610000
     PFMAL6P1_06       600001    710000
     PFMAL6P1_07       700001    810000
     PFMAL6P1_08       800001    910000
     PFMAL6P1_09       900001   1010000
     PFMAL6P1_10      1000001   1110000
     PFMAL6P1_11      1100001   1210000
     PFMAL6P1_12      1200001   1310000
     PFMAL6P1_13      1300001   1378756
Continuation (14 of 14) of PFMAL6P1 from base 1300001 (AL844505 Plasmodium falciparum 3D

  Query Match             5.1%;  Score 51.4;  DB 2;  Length 78756;
  Best Local Similarity   55.2%;  Pred. No. 0.031;
  Matches  100;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        811 AGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACC 870
              || ||  ||||| | ||||| | || |   |||  |    |||    |||    |||   |
Db       1789 AGTTTGATTATCATTTGAGAAAAAAAAAAGCACAAGGAATTGAATCAGTATTAGTACATTC 1848

Qy        871 AAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAA 930
              | |   | | ||  ||         | | ||| |  || | ||  |||||| | ||  |
Db       1849 AGATATTTTTAAATTTCTTGGTACCCATCCTAATAAATTTTTAAAATCATTAGTTGTGTA 1908

Qy        931 TCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTG 990
              | || |||||    ||||||    |  || ||||  | ||| ||||||| |     ||||
Db       1909 TGAACTTTCTAGAAAATAATTATCATTATTAGAATTTTCATCAGAAGAAAATTTTATTTC 1968

Qy        991 A 991
              |
Db       1969 A 1969


RESULT 14
PFMAL6P1_12
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL6P1 Accession AL844505
     Fragment Name      Begin     End
     PFMAL6P1_00            1    110000
     PFMAL6P1_01       100001    210000
     PFMAL6P1_02       200001    310000
     PFMAL6P1_03       300001    410000
     PFMAL6P1_04       400001    510000
     PFMAL6P1_05       500001    610000
     PFMAL6P1_06       600001    710000
     PFMAL6P1_07       700001    810000
     PFMAL6P1_08       800001    910000
     PFMAL6P1_09       900001   1010000
     PFMAL6P1_10      1000001   1110000
     PFMAL6P1_11      1100001   1210000
     PFMAL6P1_12      1200001   1310000
     PFMAL6P1_13      1300001   1378756
Continuation (13 of 14) of PFMAL6P1 from base 1200001 (AL844505 Plasmodium falciparum 3

  Query Match             5.1%;  Score 51.4;  DB 2;  Length 110000;
  Best Local Similarity   55.2%;  Pred. No. 0.031;
  Matches  100;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        811 AGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACC 870
              || ||  ||||| | ||||| | || |   |||  |    |||    |||    |||   |
Db     101789 AGTTTGATTATCATTTGAGAAAAAAAAAAGCACAAGGAATTGAATCAGTATTAGTACATTC 101848

Qy        871 AAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAA 930
              | |   | | ||  ||         | | ||| |  || | ||  |||||| | ||  |
Db     101849 AGATATTTTTAAATTTCTTGGTACCCATCCTAATAAATTTTTAAAATCATTAGTTGTGTA 101908

Qy        931 TCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTG 990
              | || |||||    ||||||    |  || ||||  | ||| ||||||| |     ||||
Db     101909 TGAACTTTCTAGAAAATAATTATCATTATTAGAATTTTCATCAGAAGAAAATTTTATTTC 101968

Qy        991 A 991
              |
Db     101969 A 101969


RESULT 15
ATT5C2
LOCUS       ATT5C2                 103098 bp    DNA     linear   PLN 17-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T5C2.
ACCESSION   AL138664
VERSION     AL138664.1  GI:6899956
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 103098)
  AUTHORS   Obermaier,B., Ottenwaelder,B., Duchemin,D., Zeitler,K., Mewes,H.W.,
            Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103098)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
```

```
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .103098
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
     gene            448. .2895
                     /gene="T5C2_10"
     gene            complement(join(448. .1062,1145. .1610,1689. .1759,
                     1855. .2375,2463. .2663,2799. .2895))
                     /gene="T5C2_10"
     CDS             complement(join(448. .1062,1145. .1610,1689. .1759,
                     1855. .2375,2463. .2663,2799. .2895))
                     /gene="T5C2_10"
                     /note="similarity to hypothetical protein At2g15420 -
                     Arabidopsis thaliana, EMBL:AC006920"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB81572.1"
                     /db_xref="GI:7263606"
                     /translation="MDCGWCDLNGYGTNQDAYRGINDGLPVDDFGGQMESREVDEDDI
                     VVYTPHRSIVVGLRSLRWTPLGDESVVVPDPRDDGSTLCTRKRLTHLIDQSIFPDDAG
                     LRFPLPDFLVRYCLRRRIAFSQLALAAVCNAVGLMMLGADREVEVDAGLLEEATTFAA
                     PKDNPGLCQVNARPNHKLVKGMKSKVPDWRLYFFFVELSEISVSDLDVLFFFASGICF
                     LVDLFSYAHVTVVNVAFLPRLVLFAAHDFQPRGTFGGISASLAAVRTRKPLHWPDIHR
                     QLQRWEKLAISFRFTLIRSGSGEIETNQQITNEDRATQLSDARVGKKASGSDSEDRAM
                     RLLDARARKIASNSVTEDGAIFREPLSVHASPRSKVSHGGSSRDSFPVVHERGSTQDI
                     REVDHSFSFNYDRKGYPFSDSRSYCGEFARKVNGLSRENPPVESPIAKNLYQTWACKE
                     VQSIATGNALYNFLEEQAREAREEMVAVEAKRKIAEFTAATRLRNLNSEKGRIEGFEK
                     QILELSESLKKSDSFGRAEQSRLRSSRDDALRTTRSERSRFKDDIALYKARIDRLRKH
                     ILDKRVADEPLFMQCQVAGSRKLLKKLITNGLAVPGETMCELKVDKNKWGKVVDDLAI
                     TDLLDDDLVAFPEVEDEAAMTTAEGEARIGGT"
     exon            complement(448. .1062)
                     /gene="T5C2_10"
                     /number=1
     intron          complement(1063. .1144)
                     /gene="T5C2_10"
                     /number=1
     exon            complement(1145. .1610)
                     /gene="T5C2_10"
                     /number=2
     intron          complement(1611. .1688)
                     /gene="T5C2_10"
                     /number=2
     exon            complement(1689. .1759)
                     /gene="T5C2_10"
                     /number=3
     intron          complement(1760. .1854)
                     /gene="T5C2_10"
                     /number=3
     exon            complement(1855. .2375)
                     /gene="T5C2_10"
                     /number=4
     intron          complement(2376. .2462)
                     /gene="T5C2_10"
                     /number=4
     exon            complement(2463. .2663)
                     /gene="T5C2_10"
                     /number=5
     intron          complement(2664. .2798)
                     /gene="T5C2_10"
                     /number=5
     exon            complement(2799. .2895)
                     /gene="T5C2_10"
                     /number=6
     misc_feature    6611. .8903
                     /note="homology to fragments of reverse transcriptase"
     gene            9986. .11518
                     /gene="T5C2_20"
     gene            complement(9986. .11518)
                     /gene="T5C2_20"
     CDS             complement(9986. .11518)
                     /gene="T5C2_20"
                     /note="similarity to retrotransposon finger protein (clone
                     DW15) - Arabidopsis thaliana, PIR:S65811"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB81573.1"
                     /db_xref="GI:7263607"
                     /translation="MEPLVASQPPPVTGEKGLLQISDLSVVVHKQIDGPEKIVESHQK
                     EDGLGSNSDSLVLGKSILGVGESQEATSEASRGTSHVTKKSWLQAAQKRSFTKQKFEV
                     VEVDGQEKVVVPKEVFLGAKPLWEDPLVGNILNSKAPHVGKIHMIVNKIWRLGDKSTL
                     IDVLEVNESTVKFRVRNEAMRHRILNRRMWNIMGLPMIISKWTPYSEDVQPALKSIPL
                     WVTLSNVPLTMFTDKGLEFLASVVGKPVRLHPKTEACESFEEAKILVEADLTKDLPKE
                     YILSGEEEGELDAVIHYSYPWLSPHCQVCKKWGHLRDSCLLEKQKPTSTLNNSPVSEK
                     EPQKIVVEEAHVSTKVVEELFNNLPSQSGVSVALETKVQSNTKTDQGWISPLKTTKSP
                     MKNSEVLKFGEVSLLSNPYAPLSVQEDIREDEIVVTTTSEINVVGQSMVTEKAVETDF
                     QTSDISGVGGRNRCPKSSHPTRKQQPRGSKTAHKVIPQGNTHLKKEPVRDQGNKIPPH
                     HH"
     exon            complement(9986. .11518)
                     /gene="T5C2_20"
                     /number=1
     gene            13872. .16433
                     /gene="T5C2_30"
     CDS             join(13872. .13930,14081. .14166,14579. .14656,14735. .14956,
                     15045. .15084,15306. .15482,15566. .15855,15916. .16433)
                     /gene="T5C2_30"
                     /note="similarity to putative proteins - Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB81574.1"
                     /db_xref="GI:7263608"
                     /translation="MGEVSPYTAIKNLKHFKTSCSRIHIFLCYFQGNKIQAGIKKNIY
                     SNYRGYAEENGTNTDTDEDDDWSDYHNSSESDYEENLQEDFYWSSQEECSSYSSLDED
                     VNAEDMAQIPDDIIYEAKDKFDILNMLENAFGKPKSKPVIPPNPKSAYVDEGDPVYSC
                     DKGKVDHCLPKDRGPNMFAIQGENYYLMGALKPKSGAKAKFQQLYIVDTENEVKNRYN
                     IMSNENESENGQKKKKIMKLLNRVNPHVKAFRYARDRFNTDAQESFHMRIIAARKTDP
                     RIYNLPTASEVAALIPGDFHEDMDKLDIVLQETTSGKLKRINGYRLGIKKARTNGHGD
                     GKKEQKDVSMRQWFAYRLQERKAEKHILFRSKRLLQQFIVDAYTMIESNRLRYIKKNQ
                     PKFRSSTREPIQNASNSGNNDLENQGKEVKIPASFTGGPRYMTNNYLDAMAICKHFGF
                     PSLFITFTCNPKWPEIVRFCKERNLKSEDRQDIICRILR"
     exon            13872. .13930
                     /gene="T5C2_30"
                     /number=1
     intron          13931. .14080
                     /gene="T5C2_30"
                     /number=1
     exon            14081. .14166
                     /gene="T5C2_30"
                     /number=2
     intron          14167. .14578
                     /gene="T5C2_30"
                     /number=2
     exon            14579. .14656
                     /gene="T5C2_30"
                     /number=3
     intron          14657. .14734
                     /gene="T5C2_30"
                     /number=3
     exon            14735. .14956
                     /gene="T5C2_30"
                     /number=4
     intron          14957. .15044
                     /gene="T5C2_30"
                     /number=4
     exon            15045. .15084
                     /gene="T5C2_30"
                     /number=5
```

```
     intron          15085. .15305
                     /gene="T5C2_30"
                     /number=5
     exon            15306. .15482
                     /gene="T5C2_30"
                     /number=6
     intron          15483. .15565
                     /gene="T5C2_30"
                     /number=6
     exon            15566. .15855
                     /gene="T5C2_30"
                     /number=7
     intron          15856. .15915
                     /gene="T5C2_30"
                     /number=7
     exon            15916. .16433
                     /gene="T5C2_30"
                     /number=8
     repeat_unit     18563. .18592
                     /note="30 bp AT tandem repeat"
     gene            19386. .20376
                     /gene="T5C2_40"
     gene            complement(join(19386. .19445,19766. .19822,20269. .20376))
                     /gene="T5C2_40"
     CDS             complement(join(19386. .19445,19766. .19822,20269. .20376))
                     /gene="T5C2_40"
                     /note="similarity to hypothetical protein At2g39150 -
                     Arabidopsis thaliana, EMBL:AC004697"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB81575.1"
                     /db_xref="GI:7263609"
                     /translation="MAISEGTMVEGIHCVPDLVELMPKQHDIPRAWLRIVVHEGRNHE

  Query Match             5.0%;  Score 50.8;  DB 8;  Length 103098;
  Best Local Similarity   74.4%;  Pred. No. 0.044;
  Matches   64;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         36 TCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGGAGACGTCATCGAACTCAATCG 95
              | || |   | | |||| |||| |||||||| ||| |||||| |||||| || |||||||
Db      69994 TTTCTTCTCCTTCGCGTTTATCACCGGAGCAATTGTCGGAGATGTCATCAAAGTCAATCG 70053

Qy         96 ATCTCAGAGGGAGTTCGATTATTTCG 121
                | | | || || || ||||||||| |
Db      70054 CTATAACAGAGAATTTGATTATTTGG 70079


Search completed: January 21, 2004, 22:14:34
Job time : 3698.68 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 260.957 Seconds
(without alignments)
10468.497 Million cell updates/sec

Title: US-09-936-869-5
Perfect score: 1012
Sequence: 1 atcgaattaaagtcaatggc.....ttctatggtaaaaaaaaaaa 1012

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1012	100.0	1012	21	AAA99365	Plant PrAG1 promot
	2	973.8	96.2	1069	21	AAC37325	Arabidopsis thalia
	3	170.8	16.9	822	15	AAQ71177	E. pinnatifida rib
	4	83.2	8.2	266	25	ABX23625	Human GDP-mannose
	5	76.6	7.6	229	25	ABX23678	Human GDP-mannose
	6	63.8	6.3	224	25	ABX26579	Human GDP-mannose
	7	50.4	5.0	1030	21	AAC34348	Arabidopsis thalia
	8	46.8	4.6	693	24	ABZ14066	Arabidopsis thalia
	9	46.8	4.6	1026	21	AAC47291	Arabidopsis thalia
c	10	43.6	4.3	2155561	24	ABN71527	Streptococcus poly
c	11	42.6	4.2	47108	24	ABK31510	Signal transductio
	12	40.8	4.0	19513	24	ABL49339	Human polynucleoti
	13	40.8	4.0	19513	24	AAS61245	Human gene regulat
c	14	40.4	4.0	10133	24	ABL32458	Human immune syste
c	15	40.2	4.0	50000	24	ABL55643	AmEPV genome fragm
	16	40	4.0	34688	24	ABQ67060	Human angiogenesis
	17	39.6	3.9	6729	24	ABQ67153	Human angiogenesis
	18	39.4	3.9	1278	19	AAV65303	DNA encoding a S.
	19	39.4	3.9	1278	23	AAS55528	Streptococcus pneu
	20	39.4	3.9	1287	25	ABX06962	S. pneumoniae type
	21	39.4	3.9	2764	19	AAV65241	DNA encoding a S.
c	22	39.4	3.9	13121	19	AAV52259	Streptococcus pneu
c	23	39.4	3.9	2162598	25	ABS56454	Streptococcus pneu
c	24	39.2	3.9	365	22	AAI88155	Human polynucleoti
	25	39.2	3.9	381	25	ABX19331	Human GDP-mannose
	26	39.2	3.9	387	25	ABX18894	Human GDP-mannose
c	27	39	3.9	7810	22	AAS45434	Chemically pretrea
c	28	39	3.9	7810	24	ABK28281	DNA transcription
c	29	38.4	3.8	5520	24	ABL33812	Human immune syste
	30	38.4	3.8	6106	22	AAS46430	Tumour suppressor
	31	38.4	3.8	6106	24	ABK40032	Human chemically p
	32	38.4	3.8	6106	24	ABL33473	Human immune syste
	33	38.4	3.8	6123	24	ABL32820	Human immune syste
	34	38.4	3.8	6265	24	ABL33417	Human immune syste
c	35	38.4	3.8	10094	25	ABX77219	5' sequence flanki
	36	38.4	3.8	116277	20	AAX20249	Borrelia burgdorfe
c	37	38.4	3.8	910715	20	AAX20248	Borrelia burgdorfe
c	38	38	3.8	789	20	AAX24990	Staphylococcus aur
c	39	38	3.8	789	24	AAD42494	Staphylococcus aur
	40	38	3.8	1342	21	AAA26424	Human secreted pro
	41	38	3.8	6591	22	AAS46284	Tumour suppressor
c	42	38	3.8	13643	23	ABL03848	Drosophila melanog
c	43	37.8	3.7	757	24	ABQ67170	Human angiogenesis
	44	37.8	3.7	831	20	AAX30136	Streptococcus pneu
	45	37.8	3.7	1290	22	AAH90726	CFE 30 coding sequ

ALIGNMENTS

RESULT 1
AAA99365
ID AAA99365 standard; DNA; 1012 BP.
XX
AC AAA99365;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter related gene sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART-) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;
XX
DR WPI; 2000-594442/56.
DR P-PSDB; AAB26796.
XX

PT Novel plant reproductive tissue promoter, useful to produce plants
PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 15; Page 42-43; 51pp; English.
XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 1012 BP; 287 A; 197 C; 217 G; 311 T; 0 other;

Query Match 100.0%; Score 1012; DB 21; Length 1012;
Best Local Similarity 100.0%; Pred. No. 1.2e-272;
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60
Db    1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60

Qy   61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120
Db   61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120

Qy  121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180
Db  121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180

Qy  181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240
Db  181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240

Qy  241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300
Db  241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300

Qy  301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360
Db  301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360

Qy  361 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 420
Db  361 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 420

Qy  421 TGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTAC 480
Db  421 TGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTAC 480

Qy  481 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 540
Db  481 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 540

Qy  541 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 600
Db  541 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 600

Qy  601 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660
Db  601 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660

Qy  661 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 720
Db  661 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 720

Qy  721 AGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 780
Db  721 AGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 780

Qy  781 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 840
Db  781 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 840

Qy  841 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 900
Db  841 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 900

Qy  901 TATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960
Db  901 TATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960

Qy  961 AGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
Db  961 AGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
```

RESULT 2
AAC37325
ID AAC37325 standard; DNA; 1069 BP.
XX
AC AAC37325;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 16971.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

```
PR   29-OCT-1999;   99US-0162142.

  Query Match             96.2%;  Score 973.8;  DB 21;  Length 1069;
  Best Local Similarity   99.0%;  Pred. No. 5.8e-262;
  Matches 1001;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    2;

Qy          1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 78

Qy         61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 138

Qy        121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 198

Qy        181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 AACGCTTGTTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 258

Qy        241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 318

Qy        301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 378

Qy        361 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 TCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACT 438

Qy        421 TGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTAC 480
              ||||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        439 TGTTCGTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACGCTTAATCTCTAC 498

Qy        481 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 TTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAA 558

Qy        541 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAA 618

Qy        601 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAG 678

Qy        661 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTA 738

Qy        721 AGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 780
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        739 AGTTTGCCGGAATACACGCCATTAGA-GGTGAAGCTATGGTTCTGAAGATGCCAACAGAA 797

Qy        781 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACAT 857

Qy        841 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 ACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTC 917

Qy        901 TATTGG--ATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGA 958
              ||||||  |||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        918 TATTGGATATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAAAAGA 977

Qy        959 TCAGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAA 1009
              ||||||||||||||||||||||||||||||||||||||||||||||    ||
Db        978 TCAGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAATTGAA 1028
```

```
RESULT 3
AAQ71177
ID   AAQ71177 standard; cDNA; 822 BP.
XX
AC   AAQ71177;
XX
DT   25-MAR-2003  (updated)
DT   27-MAR-1995  (first entry)
XX
DE   E. pinnatifida ribonuclease P8 cDNA.
XX
KW   Ribonuclease; RNAase; protein P8; pathogenic fungus;
KW   nematode damage; crop improvement; transgenic plant;
KW   disease-resistance; ds.
XX
OS   Engelmannia pinnatifida.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..801
FT                   /*tag=  a
FT   mat_peptide     10..798
FT                   /*tag=  b
XX
PN   WO9418335-A2.
XX
PD   18-AUG-1994.
XX
PF   24-JAN-1994;   94WO-US00844.
XX
PR   29-JAN-1993;   93US-0010403.
PR   16-DEC-1993;   93US-0166172.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hironaka CM,  Huynh QK,  Shah DM;
XX
DR   WPI; 1994-279756/34.
DR   P-PSDB; AAR60474.
XX
PT   Control of plant pathogenic fungi using ribonucleases - also
PT   useful for control of nematode damage to plants
XX
PS   Disclosure; Page 28-29; 35pp; English.
XX
CC   E. pinnatifida leaf cDNA expressing ribonuclease P8 was
CC   expressed in transgenic plants to improve resistance to fungal
CC   pathogens and to nematode attack.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 822 BP; 228 A; 175 C; 173 G; 246 T; 0 other;

  Query Match             16.9%;  Score 170.8;  DB 15;  Length 822;
  Best Local Similarity   56.0%;  Pred. No. 1.8e-37;
  Matches  366;  Conservative    0;  Mismatches  282;  Indels    6;  Gaps    2;

Qy         29 TATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGGAGACGTCATCGAAC 88
              |||  |||| |||  ||   |      ||  |   |    || |  ||  || |     |
Db         41 TATCAGTTCTGCTTACCGGATTGTTCTCCGTTGATGGGAGAGCAGTTGATCTCGTTTCAA 100

Qy         89 TCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCT 148
                        || |   |||| |||   ||| || |  |||| |||||||| | |||  |
Db        101 CTCTCACCGAACACACAAAGTTTGATTTCTTCACACTGGCTCTGCAATGGCCAGCAACTT 160

Qy        149 ATTGCCGTGGAACT---CGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATG 205
                |||        |     | | || ||  |  |||| | ||| |||  |||     ||
Db        161 TCTGCTCAACGCATGAGAACAAATGCTGTCCTGAAAATGGTTGTTGCCAAGGAGGAAATT 220

Qy        206 CTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTT 265
              ||||| |   ||| ||||| || ||  ||||||| || || |  ||||   | ||||| |
Db        221 CTCCACCAGGATTTACAATCCACGGACTATGGCCCGATTACAGTGATGGAACATGGCCAT 280
```

```
Qy        266 CATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTG 325
              ||||||||     | ||||  | | | || |  |||||||| | ||| | |||| | |
Db        281 CATGTTGTCCAGGTTCTGCTTATGATGAAACAAAGATTTCACCATTGCTTGATGCTTTGC 340

Qy        326 AGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGT 385
              | |  ||   ||| |  || |||||     |  | |||||| ||     || |  ||
Db        341 ACACATATATGCCGATACTAAGTTGTAACGGTATAGAATCATGCGATAACAAGACAAAGT 400

Qy        386 CATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATG 445
              | |||||||   |  | |||||  ||||||| || |||||  |    ||        ||
Db        401 CGTTTTGGGCTTATCAATGGGAACAACATGGTACATGTTCGACCTTAGTAACAGGCGACC 460

Qy        446 AGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCC 505
              | |||   || ||| | || |  ||| |   || ||| || |||||| |||| || |
Db        461 AATATGGATATTTCTTGACAAGCCTTGAATTGTATTTGGAGTATAATGTGACGGAAGTTC 520

Qy        506 TTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAA 565
              | |   |||||||    |    |||| ||      | | ||  |    |  | || ||
Db        521 TATCAAAAGCTGGAATTAAACCTTCAAATTCAAAGACTTACTCATCTGCTGCCATTGTTT 580

Qy        566 CAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAA---GAGATGCAA 622
              | | ||| | | ||| ||   |  | ||  || |||| |||||||    |||| ||
Db        581 CTACTATTGAAACTGCTTTCGGTGCACCACCGCAATTGGTCTGCAAAAATGGTGATGTAA 640

Qy        623 TCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTG 676
              |  | ||| | ||   | || ||    ||  |   |||  | | || |||| |
Db        641 TTAAAGAAGTCCGGTTGTGTTTTACCAAGCACCACAAGATCCGAGAGTGTGTGG 694
```

RESULT 4
ABX23625
ID ABX23625 standard; cDNA; 266 BP.
XX
AC ABX23625;
XX
DT 10-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #5682.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.
XX
OS Homo sapiens.
XX
PN US2002110548-A1.
XX
PD 15-AUG-2002.
XX
PF 11-JUN-2001; 2001US-0878574.
XX
PR 22-NOV-1996; 96US-0753233.
PR 03-DEC-1997; 97US-0984246.
PR 09-SEP-1998; 98US-0149674.
PR 14-JUN-1999; 99US-0333177.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Sullivan F, Kriz R, Kumar R;
XX
DR WPI; 2003-066673/06.
XX
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT rejection -
XX
PS Disclosure; SEQ ID NO 5684; 6pp; English.
XX
CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC human GM4,6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 266 BP; 49 A; 83 C; 65 G; 69 T; 0 other;

```
  Query Match             8.2%;  Score 83.2;  DB 25;  Length 266;
  Best Local Similarity   73.2%;  Pred. No. 3.4e-13;
  Matches  120;  Conservative    0;  Mismatches   43;  Indels    1;  Gaps    1;

Qy         99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGG 158
              |||  | ||||| || |||||||| |   |  | |||||||| || || || |||     |
Db        104 TCAACGAGAGTTTGACTATTTCGCTTTCGC-GTGCAATGGCCCGGTACTTACTGCAAACG 162

Qy        159 AACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATT 218
              || |||  ||| ||| |||  |||| ||||||||||||||||| ||  ||||||    ||
Db        163 CACCCGCAGTTGCTGCCCCACCAACGGTTGCTGCAGAGGCTCCAATTTTCCAACAGTGTT 222

Qy        219 CACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGC 262
              ||| || |||||  | |||||||||||||| |||||  | ||||
Db        223 CACCATACATGGACTCTGGCCTGACTATAATGATGGAACCTGGC 266
```

RESULT 5
ABX23678
ID ABX23678 standard; cDNA; 229 BP.
XX
AC ABX23678;
XX
DT 10-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #5735.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.
XX
OS Homo sapiens.
XX
PN US2002110548-A1.
XX
PD 15-AUG-2002.
XX
PF 11-JUN-2001; 2001US-0878574.
XX
PR 22-NOV-1996; 96US-0753233.
PR 03-DEC-1997; 97US-0984246.
PR 09-SEP-1998; 98US-0149674.
PR 14-JUN-1999; 99US-0333177.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Sullivan F, Kriz R, Kumar R;
XX
DR WPI; 2003-066673/06.
XX
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)

PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT rejection -
XX
PS Disclosure; SEQ ID NO 5737; 6pp; English.
XX
CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC human GM4,6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 229 BP; 49 A; 69 C; 51 G; 60 T; 0 other;

Query Match 7.6%; Score 76.6; DB 25; Length 229;
Best Local Similarity 67.5%; Pred. No. 2.3e-11;
Matches 137; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

```
Qy         99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTA-TTGCCGTG 157
              |||  | ||||| || ||| ||||| |   |  | ||||||| || || || | ||
Db         27 TCAACGAGAGTTTGACTATATCGCTTTCGCGTTGCAATGGCCCGGTACTTACTCGCAAAC 86

Qy        158 GAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAAT 217
              | || |||  ||| ||  ||| |||| ||||||||||| ||||| ||  ||||||    |
Db         87 GCACCCGCAGTTGCTGACCCACCAACGGTTGCTGCAGAAGCTCCAATTTTCCAACAGTGT 146

Qy        218 TCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCC-TTCATGTTGTTAT 276
              |||| || |||||  | ||||||||||||| | ||| |  | |||||  | | |||||  | |
Db        147 TCACCATACATGGACTCTGGCCTGACTATTATGATAGACCCTGGCCGTACCTGTTGCTCT 206

Qy        277 CGATCTGACTTTAAAGAGAAGGA 299
               |||||   ||  |    || ||
Db        207 GGATCTAGTTTCGATCCTAAAGA 229
```

RESULT 6
ABX26579
ID ABX26579 standard; cDNA; 224 BP.
XX
AC ABX26579;
XX
DT 11-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #8636.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.
XX
OS Homo sapiens.
XX
PN US20Q2110548-A1.
XX
PD 15-AUG-2002.
XX
PF 11-JUN-2001; 2001US-0878574.
XX
PR 22-NOV-1996; 96US-0753233.
PR 03-DEC-1997; 97US-0984246.
PR 09-SEP-1998; 98US-0149674.
PR 14-JUN-1999; 99US-0333177.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Sullivan F, Kriz R, Kumar R;
XX
DR WPI; 2003-066673/06.
XX
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT rejection -
XX
PS Disclosure; SEQ ID NO 8638; 6pp; English.
XX
CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC human GM4,6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 224 BP; 37 A; 71 C; 53 G; 63 T; 0 other;

Query Match 6.3%; Score 63.8; DB 25; Length 224;
Best Local Similarity 69.9%; Pred. No. 8.5e-08;
Matches 86; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

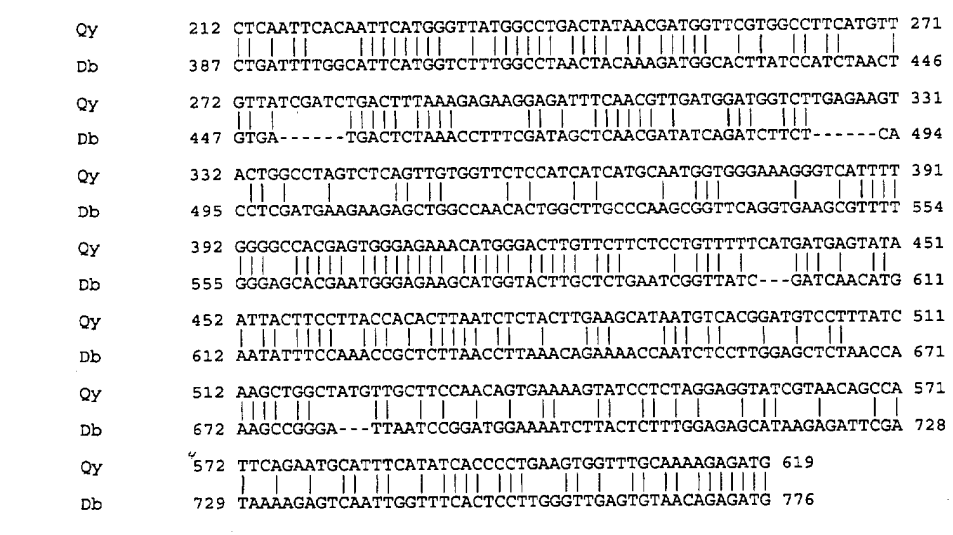
```
Qy         99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGG 158
              |||  | ||||| || |||||||| |   |  | |||||||| || || || |||    |
Db        102 TCAACGAGAGTTTGACTATTTCGCTTTCGCGTTGCAATGGCCCGGTACTTACTGCAAACG 161

Qy        159 AACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATT 218
               || |||  ||  ||| |||  | || ||||||||||||||| ||  ||||||    ||
Db        162 CACCCGCAGTTTCTGCCCCATCATCGGTTGCTGCAGAGGCTCCAATTTTCCAACAGTGTT 221

Qy        219 CAC 221
              |||
Db        222 CAC 224
```

RESULT 7
AAC34348
ID AAC34348 standard; DNA; 1030 BP.
XX
AC AAC34348;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6327.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 50.4; DB 21; Length 1030;
Best Local Similarity 52.5%; Pred. No. 0.00091;
Matches 214; Conservative 0; Mismatches 176; Indels 18; Gaps 4;



RESULT 8
ABZ14066

ID ABZ14066 standard; DNA; 693 BP.
XX
AC ABZ14066;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1871.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT -
XX
PS Claim 144; SEQ ID NO 1871; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 693 BP; 182 A; 150 C; 145 G; 216 T; 0 other;

Query Match 4.6%; Score 46.8; DB 24; Length 693;
Best Local Similarity 56.0%; Pred. No. 0.0078;
Matches 131; Conservative 0; Mismatches 97; Indels 6; Gaps 2;

RESULT 9
AAC47291
ID AAC47291 standard; DNA; 1026 BP.
XX
AC AAC47291;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 53278.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.6%; Score 46.8; DB 21; Length 1026;
Best Local Similarity 56.0%; Pred. No. 0.0092;
Matches 131; Conservative 0; Mismatches 97; Indels 6; Gaps 2;

```
Qy        386 CATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATG 445
              | |||||||  ||||| |||||||| ||||| ||||| |||    | ||| |    |||
Db        549 CGTTTTGGGAGCACGAATGGGAGAAGCATGGTACTTGCTCTGAATCGGTTAT---CGATC 605

Qy        446 AGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCC 505
              |  || | || |||| ||| | ||||| ||  |   |||    ||| ||    |  | |
Db        606 AACATGAATATTTCCAAACCGCTCTTAACCTTAAACAGAAAACCAATCTCCTTGGAGCTC 665

Qy        506 TTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAA 565
              |    |||| ||     ||  | |  |  |  ||   ||  || | |   | ||   |
Db        666 TAACCAAAGCCGGGA---TTAATCCGGATGGAAAATCTTACTCTTTGGAGAGCATAAGAG 722
```

```
Qy        566 CAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATG 619
              | ||  |  |  || ||  | |||| |||   || |  || || |||||||
Db        723 ATTCGATAAAAGAGTCAATTGGTTTCACTCCTTGGGTTGAGTGTAACAGAGATG 776
```

RESULT 10
ABN71527/c
ID ABN71527 standard; DNA; 2155561 BP.
XX
AC ABN71527;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 10967.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus sp.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 8; Page 4196-4488; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 4.3%; Score 43.6; DB 24; Length 2155561;
Best Local Similarity 55.2%; Pred. No. 1.9;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy        856 TGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATC 915
              | ||| |  |   | | ||  |  |||  ||    | ||   |||||||| |||| ||||
Db    1665003 TATAAGTGAAAATCTAGAAATTTCAAATTTTTCAATTTTATTTTCTATTGTATATTTATC 1664944
```

```
Qy        916 ATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGA 975
              |||| |  || |  || | |      | |||  |  ||| | |||||   || ||
Db    1664943 TTTATTGATACTAGTTTCAGGATTTACAGTAAAATTGATGAGAAGAAAAAATGATTATGA 1664884

Qy        976 AGATATCAACATTTGAGTTCTATGGTAAAAAAAA 1009
              ||| | | ||||   | |  || |||||||
Db    1664883 AGAGCTAGAAATTTTCCCTGGACAGTTAAAAAAA 1664850
```

RESULT 11
ABK31510/c
ID ABK31510 standard; DNA; 47108 BP.
XX
AC ABK31510;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #177.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07472.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
PS Claim 1; SEQ ID No 353; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 47108 BP; 13283 A; 429 C; 9740 G; 23656 T; 0 other;

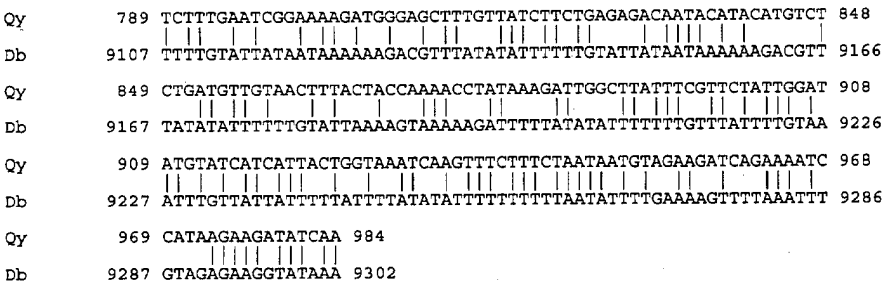
Query Match 4.2%; Score 42.6; DB 24; Length 47108;
Best Local Similarity 55.8%; Pred. No. 0.7;
Matches 101; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

```
Qy        833 CAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTA 892
              |||||| |||||      | || || |  || ||| ||    |   | || |       ||
Db       7217 CAATACTTACATAAACATTATTTTCTTCTTTAACTTCCCCCAAAAAACAAAACCAAAATA 7158

Qy        893 TTTCGTTCTATTGGATATGTATCATCATTAC-TGGTAAATCAAGTTTCTTTCTAATAATG 951
                  | ||   |||| ||| |  |  || |  |||| | | ||    || |||||
Db       7157 ACATTACCCATAACATATATATAAATAAAACATAATAAAACTAATTATAATCAAATAACC 7098

Qy        952 TAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAA 1011
              || || ||||  ||| ||| || || ||||  | |||| | ||   |  |  |||| |||
Db       7097 TAAAAAATCACGAAAACCAAAAAAATATATTTATATTTTATTTTATTTTTTTAAAACAAA 7038

Qy       1012 A 1012
              |
Db       7037 A 7037
```

RESULT 12
ABL49339
ID ABL49339 standard; DNA; 19513 BP.
XX
AC ABL49339;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human polynucleotide associated with DNA replication SEQ ID NO 39.
XX
KW Human; cytostatic; neuroprotective; nootropic; immunostimulant;
KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW ERCC3; SNRPA1; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177377-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03971.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017471/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with DNA replication, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia -
XX
PS Claim 1; SEQ ID NO 39; 23pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acid sequences comprising at least 18
CC bases of a chemically pretreated gene associated with gene regulation,
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPA1,
CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
CC unmethylated at the 5-position to uracil or another base with
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
CC cytosine methylations. The DNA sequences and method are useful in the
CC diagnosis of diseases (or predisposition to diseases) associated with DNA
CC replication and in therapy of such diseases, by enabling analysis of the
CC cytosine methylation patterns of such genes. They are especially useful
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
CC syndrome, solid tumours and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification and was supposed to be available directly from WIPO at

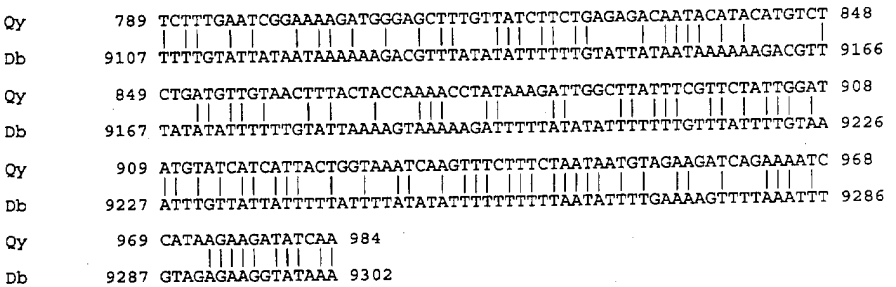
CC ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
CC not correspond to that referred to in the specification. The present data
CC is taken from EPO data for the patent.
XX
SQ Sequence 19513 BP; 4543 A; 563 C; 5296 G; 9111 T; 0 other;

Query Match 4.0%; Score 40.8; DB 24; Length 19513;
Best Local Similarity 50.5%; Pred. No. 1.5;
Matches 99; Conservative 0; Mismatches 97; Indels 0; Gaps 0;



RESULT 13
AAS61245
ID AAS61245 standard; DNA; 19513 BP.
XX
AC AAS61245;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #200.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 205; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 19513 BP; 4543 A; 563 C; 5296 G; 9111 T; 0 other;

Query Match 4.0%; Score 40.8; DB 24; Length 19513;
Best Local Similarity 50.5%; Pred. No. 1.5;
Matches 99; Conservative 0; Mismatches 97; Indels 0; Gaps 0;



RESULT 14
ABL32458/c
ID ABL32458 standard; DNA; 10133 BP.
XX
AC ABL32458;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 431.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX

PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 431; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 10133 BP; 3075 A; 70 C; 2051 G; 4937 T; 0 other;

Query Match 4.0%; Score 40.4; DB 24; Length 10133;
Best Local Similarity 52.4%; Pred. No. 1.5;
Matches 89; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 843 ATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTA 902
Db 2924 ATAACAATTACGTACAATATATAATTACAAATTATATATATATATATATTATAAAAACAA 2865

Qy 903 TTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAG 962
Db 2864 CTACAAAAATATAATAACTATTATAAAAAAAACATACTAACTTAAAACCAAAAATTCAAA 2805

Qy 963 AAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
Db 2804 ACAAACCTAAACAACATAACAAAACCTTATCTCTACTAAAAAAAAAAAA 2755

RESULT 15
ABL55643/c
ID ABL55643 standard; DNA; 50000 BP.
XX
AC ABL55643;
XX
DT 01-JUL-2002 (first entry)
XX
DE AmEPV genome fragment#1.
XX
KW AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW genetic deficiency disorder; ds.
XX
OS Amsacta moorei entomopoxvirus.
XX
PN WO200212526-A2.
XX
PD 14-FEB-2002.
XX
PF 10-AUG-2001; 2001WO-US25287.
XX
PR 10-AUG-2000; 2000US-224479P.
PR 14-SEP-2000; 2000US-0662254.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Moyer RW, Li Y, Bawden AL;
XX
DR WPI; 2002-227161/28.
XX
PT Novel recombinant entomopox virus vector useful for delivering
PT polynucleotide encoding protein to vertebrate cell, comprises
PT polynucleotide encoding protein operably linked with heterologous
PT promoter sequence -
XX
PS Disclosure; Page 125-150; 326pp; English.
XX
CC The invention relates to a recombinant entomopox virus (EPV) vector,
CC comprising a polynucleotide encoding a protein operably linked with a
CC heterologous promoter sequence. The invention also concerns methods for
CC providing gene therapy for genetic deficiency disorders. Vectors of the
CC invention are useful for delivering a polynucleotide encoding a protein
CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC The vector is introduced into the vertebrate cell by infection in a viral
CC particle, or by transfection, transduction, or injection either in vitro
CC or in vivo. The vector is useful for the delivery and expression of
CC biologically useful proteins in gene therapy protocols, and for
CC delivering large DNA segments for engineering of vertebrate cells.
CC Polynucleotides of the invention have applications in techniques such as
CC their use as insertion sites for foreign genes of interest, hybridisation
CC probes, for chromosome and gene mapping, in PCR technologies, and in the
CC production of sense or antisense nucleic acids. Vectors of the invention
CC provide for stable integration and expression of heterologous DNA in host
CC cells, and are adapted for accepting large heterologous polynucleotide
CC inserts which can be delivered in an infected or transformed cell and
CC expressed in a stable fraction. The current sequence represents a
CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC (AmEPV).
XX
SQ Sequence 50000 BP; 19857 A; 4373 C; 4222 G; 21548 T; 0 other;

Query Match 4.0%; Score 40.2; DB 24; Length 50000;
Best Local Similarity 51.4%; Pred. No. 3.4;
Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 823 TTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAA 882
Db 30790 TTATTATAGAGGCTATATAAATTATTTAAAAAATTTATCGTTACAATTATAATATTTTAT 30731

Qy 883 GATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTT 942
Db 30730 AATAATAATTTTTTTTGTTATATAATTTTTATTATACATATTTTTATATGTAGTATATCT 30671

Qy 943 CTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTA 1002
Db 30670 GTATTTTTTTATATAATCTGAAATATTATGACTAGTAAACATGATTATTATATTATAATT 30611

Qy 1003 A 1003
Db 30610 A 30610

Search completed: January 21, 2004, 17:31:58
Job time : 267.957 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 340.379 Seconds
(without alignments)
10468.497 Million cell updates/sec

Title: US-09-936-869-1
Perfect score: 1320
Sequence: 1 aaactcgacagcaaatatga.....tgttgctgttgcgcccattg 1320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1320	100.0	1320	21	AAA99362	Plant PrAG1 promot
	2	1320	100.0	1401	21	AAA99363	Plant PrAG1 promot
c	3	46.6	3.5	540	24	ABQ18618	Oligonucleotide fo
	4	46.6	3.5	540	24	ABQ18619	Oligonucleotide fo
c	5	46.4	3.5	4492	23	ABL15002	Drosophila melanog
c	6	45.6	3.5	563	23	ABV50724	Human prostate exp
c	7	45	3.4	519	25	ABX22233	Human GDP-mannose
	8	44.4	3.4	15923	24	ABL70179	Chemically treated
	9	44.4	3.4	15923	24	AAS61132	Human gene regulat
	10	44.4	3.4	15923	24	ABK31220	Signal transductio
	11	44	3.3	6261	22	AAS46643	Tumour suppressor
	12	44	3.3	6261	24	AAS61046	Human gene regulat
	13	44	3.3	6261	24	ABK31160	Signal transductio
c	14	43.6	3.3	11084	20	AAX22302	Human IL-1ra BAC c
	15	43.4	3.3	409	22	AAI92030	Human polynucleoti
	16	43.4	3.3	13326	24	ABL33712	Human immune syste
c	17	43.2	3.3	18434	24	ABL34006	Human immune syste
c	18	43	3.3	201	22	ABA48463	Human breast cell
c	19	43	3.3	201	22	ABA51474	Human breast cell
c	20	43	3.3	201	22	ABA66368	Human foetal liver
c	21	43	3.3	201	22	ABA69502	Human foetal liver
c	22	43	3.3	201	22	ABA33430	Probe #11896 for g
c	23	43	3.3	201	22	ABA36424	Probe #14890 for g
c	24	43	3.3	201	22	AAK14785	Human brain expres
c	25	43	3.3	201	22	AAK17754	Human brain expres
c	26	43	3.3	201	22	AAK40526	Human bone marrow
c	27	43	3.3	201	22	AAK43573	Human bone marrow
c	28	43	3.3	201	22	AAI21285	Probe #11218 for g
c	29	43	3.3	201	22	AAI24360	Probe #14293 for g
c	30	43	3.3	201	22	AAI46560	Probe #15246 used
c	31	43	3.3	201	22	AAI49630	Probe #18316 used
c	32	43	3.3	201	22	AAI06989	Probe #6980 used t
c	33	43	3.3	201	22	AAI09899	Probe #9890 used t
c	34	43	3.3	201	23	ABS40089	Human liver single
c	35	43	3.3	201	23	ABS43218	Human liver single
c	36	43	3.3	201	24	ABS17726	Human genome-deriv
c	37	43	3.3	395	22	ABA46372	Human breast cell
c	38	43	3.3	395	22	ABA56938	Human foetal liver
c	39	43	3.3	395	22	ABA26550	Probe #5016 for ge
c	40	43	3.3	395	22	AAK05033	Human brain expres
c	41	43	3.3	395	22	AAK30563	Human bone marrow
c	42	43	3.3	395	22	AAI15180	Probe #5113 for ge
c	43	43	3.3	395	22	AAI36508	Probe #5194 used t
c	44	43	3.3	395	22	AAI04920	Probe #4911 used t
c	45	43	3.3	395	23	ABS30228	Human liver single

ALIGNMENTS

RESULT 1
AAA99362
ID AAA99362 standard; DNA; 1320 BP.
XX
AC AAA99362;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter DNA sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Pinus radiata.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;
XX
DR WPI; 2000-594442/56.
DR P-PSDB; AAB26795.
XX

PT Novel plant reproductive tissue promoter, useful to produce plants
PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 2; Page 37-38; 51pp; English.
XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 1320 BP; 390 A; 269 C; 280 G; 381 T; 0 other;

Query Match 100.0%; Score 1320; DB 21; Length 1320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60

Qy         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120

Qy        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180

Qy        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240

Qy        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300

Qy        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360

Qy        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420

Qy        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480

Qy        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540

Qy        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600

Qy        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660

Qy        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720

Qy        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780

Qy        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840

Qy        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900

Qy        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960

Qy        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020

Qy       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080

Qy       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140

Qy       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200

Qy       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260

Qy       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
```

RESULT 2
AAA99363
ID AAA99363 standard; DNA; 1401 BP.
XX
AC AAA99363;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter DNA sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Pinus radiata.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART-) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;
XX
DR WPI; 2000-594442/56.
DR P-PSDB; AAB26795.
XX
PT Novel plant reproductive tissue promoter, useful to produce plants
PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 3; Fig 2; 51pp; English.

XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 1401 BP; 406 A; 281 C; 302 G; 412 T; 0 other;

Query Match 100.0%; Score 1320; DB 21; Length 1401;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60

Qy         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120

Qy        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180

Qy        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240

Qy        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300

Qy        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360

Qy        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420

Qy        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480

Qy        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540

Qy        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600

Qy        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660

Qy        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720

Qy        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780

Qy        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840

Qy        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900

Qy        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960

Qy        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020

Qy       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080

Qy       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140

Qy       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200

Qy       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260

Qy       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
```

RESULT 3
ABQ18618/c
ID ABQ18618 standard; DNA; 540 BP.
XX
AC ABQ18618;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 5209.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 540 BP; 63 A; 40 C; 97 G; 340 T; 0 other;

Query Match 3.5%; Score 46.6; DB 24; Length 540;
Best Local Similarity 53.6%; Pred. No. 0.025;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 255 TGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTG 314
Db 534 TAAAAAAAAAACAAATAATAATAAAAAAAAAAAATAAAAAAAAAAAAAAAATAAAATAATA 475

Qy 315 GAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAA 374
Db 474 AAAAAAAAAAACGAATAAAAAAAAAAAACGATAAAAAAAAACTAAAAAAACCGAAAAAAA 415

Qy 375 GGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAA 434
Db 414 AAACGAAATAATAAAAAAAAACTAAATCAAAAAAACGAAAAAAACGATCGAAAAAACGAC 355

Qy 435 G 435
Db 354 G 354

RESULT 4
ABQ18619
ID ABQ18619 standard; DNA; 540 BP.
XX
AC ABQ18619;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 5210.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 540 BP; 340 A; 97 C; 40 G; 63 T; 0 other;

Query Match 3.5%; Score 46.6; DB 24; Length 540;
Best Local Similarity 53.6%; Pred. No. 0.025;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 255 TGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTG 314
Db 7 TAAAAAAAAAACAAATAATAATAAAAAAAAAAAATAAAAAAAAAAAAAAAATAAAATAATA 66

Qy 315 GAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAA 374
Db 67 AAAAAAAAAAACGAATAAAAAAAAAAAACGATAAAAAAAAACTAAAAAAACCGAAAAAAA 126

Qy 375 GGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAA 434
Db 127 AAACGAAATAATAAAAAAAAACTAAATCAAAAAAACGAAAAAAACGATCGAAAAAACGAC 186

Qy 435 G 435
Db 187 G 187

RESULT 5
ABL15002/c
ID ABL15002 standard; cDNA; 4492 BP.
XX
AC ABL15002;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39488.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB70899.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 39488; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4492 BP; 1158 A; 1245 C; 1141 G; 948 T; 0 other;

Query Match 3.5%; Score 46.4; DB 23; Length 4492;
Best Local Similarity 49.6%; Pred. No. 0.068;
Matches 119; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 804 TCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCA 863
Db 2878 TCCAGCGGCATGCTATCAATCTCATCGCCAACCACAACGATGCCATCGGGATCATCGTCA 2819

Qy 864 TTACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCA 923
Db 2818 TCTTCGTCCTCATCGCCGTCAGCCTCGTCGTCATCATCGTCGCCGGCCATGTGGATGCCA 2759

Qy 924 ATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAA 983
Db 2758 ACCTGCGGAAGAGCAACGGAAAATGTTATTGAATAATATGTTGCAAACTATAAACAAAGT 2699

Qy 984 AACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCAT 1043
Db 2698 TAAAAAGCAAACAACCAAGCTTTAAAATATTAAGAATATGCATGTACTATGTTTTATCAT 2639

RESULT 6
ABV50724/c
ID ABV50724 standard; cDNA; 563 BP.
XX
AC ABV50724;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 50715.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 9866; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 563 BP; 45 A; 107 C; 22 G; 385 T; 4 other;

Query Match 3.5%; Score 45.6; DB 23; Length 563;
Best Local Similarity 49.2%; Pred. No. 0.047;
Matches 120; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 267 AATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACG 326
Db 520 AAGAAAAAAAAAAAAAAAAAAGGGAAAAGAAAAAAAAAAAAAAAGAGAAAAAAAAAAAAAAG 461

Qy 327 GTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACA 386
Db 460 GACAGGAAAAAGGAAAAAAAAAAGAAGAAAAAAAACAAGAAGAAGAAAAAAAAGAAAAAAA 401

Qy 387 ACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAG 446
Db 400 AAAAAAAAAAAAAAAAAAAAAAGAAGGGAAAAAAAGAAAAAAAAAAAAAAAAAGAAAAAAA 341

Qy 447 TGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGC 506
Db 340 AAAAAAAAAAAAAAAAAGAAAGAAGAAAAAAAGAGAAGAAAAAAAAAAGAAAAAAAAAAA 281

Qy 507 AAGA 510
Db 280 AAAA 277

RESULT 7
ABX22233/c
ID ABX22233 standard; cDNA; 519 BP.
XX
AC ABX22233;
XX
DT 10-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #4290.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;

```
KW   arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW   complex carbohydrate; gene replacement therapy; immunosuppressive;
KW   antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW   antiasthmatic; vasotropic.
XX
OS   Homo sapiens.
XX
PN   US2002110548-A1.
XX
PD   15-AUG-2002.
XX
PF   11-JUN-2001; 2001US-0878574.
XX
PR   22-NOV-1996;   96US-0753233.
PR   03-DEC-1997;   97US-0984246.
PR   09-SEP-1998;   98US-0149674.
PR   14-JUN-1999;   99US-0333177.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sullivan F,  Kriz R,  Kumar R;
XX
DR   WPI; 2003-066673/06.
XX
PT   New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT   peptide, for manufacturing complex carbohydrates, or as targets for
PT   screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT   rejection -
XX
PS   Disclosure; SEQ ID NO 4292; 6pp; English.
XX
CC   The invention relates to a composition comprising a human GDP-mannose
CC   4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC   GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC   in a mammalian subject and for treating or ameliorating diseases affected
CC   by the level of cellular fucosylation or diseases affected by the
CC   fucosylation of glycoconjugates. These diseases include arthritis,
CC   transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC   infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC   useful for manufacturing complex carbohydrates and as targets for
CC   screening small molecule antagonists of the activity of the enzyme. The
CC   polynucleotide is useful in developing an assay for defects in the
CC   enzyme, as well as in gene replacement therapy. Sequences
CC   ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC   human GM4,6D peptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 519 BP; 129 A; 53 C; 23 G; 313 T; 1 other;

  Query Match             3.4%;  Score 45;  DB 25;  Length 519;
  Best Local Similarity   47.7%;  Pred. No. 0.066;
  Matches  132;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
              | |||||| | ||| ||| || ||||||||||    ||||  ||    |  ||   |||||
Db        511 AAAATTTAAAGTAAAAATAAAATATAAAAAATATGTAAATAAAAAAAATAAGATAAGAAAA 452

Qy        325 CGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCA 384
              | |       ||  | | |     |   | |  || ||      |   |||  | | | |
Db        451 AAGAGAAAACAAATAAAAAAGTATAATCAATTAAAAAAAAAACAAGTAAAGATAAAATAA 392

Qy        385 CAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGG 444
              ||  | |           |  || |  | ||||   ||   | |||  || |  | | |||
Db        391 AAAAGTTAAAAAATATAATATAAAAAAAAATAAATTTGAATAAAAAAAATGATAAAATGA 332

Qy        445 AGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAAT 504
              | |  |    || ||  ||  | |     | || |   |  |  |||| ||  | ||  |
Db        331 ATTTAAGAAAAATATATATATAAAATAAAAAATAATAATAATAAATATATATATTAAGAA 272

Qy        505 GCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTAT 541
                || | |  | |     | ||||   |  ||  |||
```

```
Db        271 TAAAAATTATAATGGATAAATATTGAAAAGAAAATAT 235


RESULT 8
ABL70179
ID   ABL70179 standard; DNA; 15923 BP.
XX
AC   ABL70179;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence#35.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease;
KW   cancer; tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP07471.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signalling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signalling -
XX
PS   Claim 1; SEQ ID NO 69; 24pp+sequence listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

  Query Match             3.4%;  Score 44.4;  DB 24;  Length 15923;
  Best Local Similarity   50.8%;  Pred. No. 0.4;
  Matches  133;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy         97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
              | | | ||| |  | ||| ||| | |     |  |  |||  |||    |  || |  |||
Db      10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

Qy        157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
              || |       | |||| | | | |  |||||| | |||     | || |||||  ||
Db      10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy        217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
              ||   ||  | |     | ||     ||   || || |                |||||  |
Db      10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945
```

```
Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
              |||| || | | | |||||| |||| ||  || |||   ||||||||    ||  |  | |
Db      10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy        337 AATATACACAGAAGAGTACTTT 358
              || |   ||     |  | |||
Db      11006 AAAAATTACGTTGTATGATTTT 11027
```

RESULT 9
AAS61132
ID AAS61132 standard; DNA; 15923 BP.
XX
AC AAS61132;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #87.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 89; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

```
  Query Match             3.4%;  Score 44.4;  DB 24;  Length 15923;
  Best Local Similarity   50.8%;  Pred. No. 0.4;
  Matches  133;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy         97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
              | | | ||| |  | ||| ||| | |   |  | |||  |||    || | |||
Db      10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

Qy        157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
              || |       | ||| | | | |  ||||| | |||     | || |||||  ||
Db      10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy        217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
              ||   ||  | |     | ||    ||   || ||| |                |||||  |
Db      10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945

Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
              |||| || | | | |||||| |||| ||  || |||   ||||||||    ||  |  | |
Db      10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy        337 AATATACACAGAAGAGTACTTT 358
              || |   ||     |  | |||
Db      11006 AAAAATTACGTTGTATGATTTT 11027
```

RESULT 10
ABK31220
ID ABK31220 standard; DNA; 15923 BP.
XX
AC ABK31220;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #32.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07472.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
PS Claim 1; SEQ ID No 63; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these

CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

Query Match 3.4%; Score 44.4; DB 24; Length 15923;
Best Local Similarity 50.8%; Pred. No. 0.4;
Matches 133; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

Qy 97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
Db 10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

Qy 157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
Db 10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy 217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
Db 10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945

Qy 277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
Db 10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy 337 AATATACACAGAAGAGTACTTT 358
Db 11006 AAAAATTACGTTGTATGATTTT 11027

RESULT 11
AAS46643
ID AAS46643 standard; DNA; 6261 BP.
XX
AC AAS46643;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #365.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX

PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
PS Claim 1; SEQ ID No 365; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

Query Match 3.3%; Score 44; DB 22; Length 6261;
Best Local Similarity 50.5%; Pred. No. 0.35;
Matches 107; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
Db 1216 GGTCGTTTGTAGTTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy 206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
Db 1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGGATAGAGCGA 1335

Qy 266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
Db 1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy 326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
Db 1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427

RESULT 12
AAS61046
ID AAS61046 standard; DNA; 6261 BP.
XX
AC AAS61046;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #1.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;

```
KW   immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW   nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200177375-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03968.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017470/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated
PT   with gene regulation, useful for analysing cytosine methylations for
PT   diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT   disease -
XX
PS   Claim 1; SEQ ID No 1; 26pp; English.
XX
CC   The invention relates to 224 nucleic acid sequences comprising at least
CC   18 bases of a chemically pretreated gene associated with gene regulation
CC   selected from 43 known genes (or complementary sequences). The
CC   chemical pretreatment converts cytosine bases unmethylated at the
CC   5-position to uracil or another base with hybridisation behaviour
CC   dissimilar to cytosine, to enable analysis of cytosine methylations.
CC   The DNA sequences, oligomers (or sets/arrays) and method are
CC   useful in the diagnosis of diseases (or predisposition to diseases)
CC   associated with gene regulation and in therapy of such diseases, by
CC   enabling analysis of the cytosine methylation patterns of such genes,
CC   kits are provided. They are especially useful in diagnosis
CC   and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC   disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC   asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC   preeclampsia, graft versus-host disease. The present sequence is a
CC   sequence included in the sequence data for this specification and is
CC   associated with the human gene regulation-associated genes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

  Query Match             3.3%;  Score 44;  DB 24;  Length 6261;
  Best Local Similarity   50.5%;  Pred. No. 0.35;
  Matches  107;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
              ||| |  |    ||  |        |||    |   | |  || ||| |         | 
Db       1216 GGTCGTTTGTAGTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy        206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
                 | |  ||| | ||  |       |  ||       ||  ||| | | || |   | |
Db       1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGATAGAGCGA 1335

Qy        266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
               | ||||||  ||| |   |||||||||||||||||| | ||  ||  ||| ||  |   
Db       1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy        326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
               |    | | |  |    ||  |||| |  ||
Db       1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427
```

```
RESULT 13
ABK31160
ID   ABK31160 standard; DNA; 6261 BP.
XX
AC   ABK31160;
XX
DT   23-APR-2002  (first entry)
XX
DE   Signal transduction associated gene modified DNA #2.
XX
KW   Human; signal transduction associated gene; cytosine methylation state;
KW   CpG island; signal transduction associated disease; solid tumour; cancer;
KW   antitumour; cytostatic; mutant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200926-A2.
XX
PD   03-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP07472.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-147896/19.
XX
PT   Oligonucleotide for diagnosis and therapy of diseases associated with
PT   signal transduction e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with signal transduction  -
XX
PS   Claim 1; SEQ ID No 3; 24pp; English.
XX
CC   The present invention relates to chemically modified DNA sequences of
CC   signal transduction associated genes. The DNA sequences are chemically
CC   modified using a solution of bisulphite, hydrogen sulphite or
CC   disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC   for detecting the cytosine methylation state (CpG islands) of these
CC   genes, and a method for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with signal transduction.
CC   The genomic DNA can be obtained from cells or cellular components which
CC   contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC   cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC   eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC   histologic object slides, and all their possible combinations. The
CC   sequences of the invention are useful for the diagnosis and therapy of
CC   diseases associated with signal transduction e.g. solid tumours and
CC   cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC   sequences of different genes associated with signal transduction, or
CC   their complementary sequences.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

  Query Match             3.3%;  Score 44;  DB 24;  Length 6261;
  Best Local Similarity   50.5%;  Pred. No. 0.35;
  Matches  107;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
              ||| |  |    ||  |        |||    |   | |  || ||| |         | 
Db       1216 GGTCGTTTGTAGTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy        206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
                 | |  ||| | ||  |       |  ||       ||  ||| | | || |   | |
Db       1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGATAGAGCGA 1335
```

```
Qy        266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
              | ||||||  ||| ||| ||||||||||||||||||| | ||  ||  ||| ||  |
Db       1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy        326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
              |    | | |  |    ||  |||| |  ||
Db       1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427


RESULT 14
AAX22302/c
ID   AAX22302 standard; DNA; 11084 BP.
XX
AC   AAX22302;
XX
DT   22-JUN-1999  (first entry)
XX
DE   Human IL-1ra BAC contiguous DNA sequence 95.
XX
KW   Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW   interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW   chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW   growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX
OS   Homo sapiens.
XX
PN   WO9906426-A1.
XX
PD   11-FEB-1999.
XX
PF   03-AUG-1998;   98WO-US16102.
XX
PR   02-JUL-1998;   98US-0091650.
PR   04-AUG-1997;   97US-0054646.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Pan Y;
XX
DR   WPI; 1999-153692/13.
XX
PT   New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT   used to inhibit inflammation and to screen for specific modulators
XX
PS   Example 5; Figure 4; 226pp; English.
XX
CC   AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
CC   sequences containing alternatively spliced forms of human IL-1ra. Such
CC   fragments are used in the method of the invention which describes the
CC   isolation of a novel human TANGO-77 encoding nucleic acid and protein.
CC   Tango-77 is a member of the cytokine superfamily that is expected to
CC   inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
CC   may also bind to a new receptor so could regulate other cellular
CC   processes associated with acute or chronic inflammation, e.g. asthma,
CC   chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and
CC   inflammatory bowel disease. It may also induce or suppress interleukins,
CC   cytokines and growth factors. Modulators of this protein are used to
CC   treat or prevent conditions associated with abnormal levels of
CC   inflammation, or activity of IL-1 or its receptor complex.
XX
SQ   Sequence 11084 BP; 3205 A; 2388 C; 2598 G; 2892 T; 1 other;

  Query Match             3.3%;  Score 43.6;  DB 20;  Length 11084;
  Best Local Similarity   48.8%;  Pred. No. 0.56;
  Matches  118;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
              |||| | | | |||| ||| ||  ||| ||   |||          | |||  |  |
Db        780 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAGATGGAAGGAAGGAAAGAGAAGAAA 721

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
                || || | | | |  | |    || || |    |  |  | |  | ||  ||
```

```
Db        720 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 661

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
                 ||    |||   |||| || |  || | |||  |||| |||| | |     |
Db        660 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 601

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
              | ||| ||| | ||   | |   || | |||| |||  |   | | | | || | |||| |
Db        600 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 541

Qy        514 CA 515
              |
Db        540 AA 539


RESULT 15
AAI92030
ID   AAI92030 standard; cDNA; 409 BP.
XX
AC   AAI92030;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 12090.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO12099.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 12090; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 409 BP; 183 A; 50 C; 77 G; 97 T; 2 other;

  Query Match             3.3%;  Score 43.4;  DB 22;  Length 409;
  Best Local Similarity   52.9%;  Pred. No. 0.16;
```

```
  Matches   92;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        151 GCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGT 210
              | |||| ||  || ||  | ||||| ||| |||  | |     || |  |||| | ||
Db        164 GTTTGGGTTTGGGTTAGAAGGAGAATGAAATAAGAGAGATTGTATTTAAACTAGCCTAGG 223

Qy        211 ACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATT 270
              | | |||    |      |   || |      |     | |   ||| ||||  | ||
Db        224 AAAATGGTGCACTGAAATAAAGGTTACTGCAGGGGTTTTCCCCATGACAAGCANAAAAAA 283

Qy        271 TATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
               | || ||| | | |||| |||||| |||||||   ||    |  ||  ||||
Db        284 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAGAAAAA 337

Search completed: January 21, 2004, 17:31:46
Job time : 344.379 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:02:36 ; Search time 2632.89 Seconds
(without alignments)
12185.050 Million cell updates/sec

Title: US-09-936-869-1
Perfect score: 1320
Sequence: 1 aaactcgacagcaaatatga.....tgttgctgttgcgcccattg 1320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		154	11.7	526	9	AW754659	AW754659 PC04G10 P
c 2		61.4	4.7	1101	29	CNS0039G	AL063921 Drosophil
3		57.2	4.3	987	29	CNS00418	AL066537 Drosophil
4		56.2	4.3	1201	13	BX461310	BX461310 BX461310
5		55.2	4.2	1101	29	CNS003B2	AL064080 Drosophil
c 6		55.2	4.2	1101	29	CNS00DG1	AL069971 Drosophil
c 7		54.8	4.2	1041	9	AL559034	AL559034 AL559034
8		54	4.1	964	29	CNS006N9	AL065781 Drosophil
9		54	4.1	1101	29	CNS0174K	AL107582 Drosophil
c 10		-54	4.1	1101	29	CNS017KE	AL108152 Drosophil
11		53.4	4.0	1101	29	CNS0100X	AL098379 Drosophil
12		52.8	4.0	997	29	CNS005TE	AL060767 Drosophil
13		52.6	4.0	634	29	BX122696	BX122696 Danio rer
14		52.6	4.0	1101	29	CNS0182P	AL108811 Drosophil
c 15		52.6	4.0	1114	13	BX407949	BX407949 BX407949
c 16		52.6	4.0	1201	9	AL566565	AL566565 AL566565
17		52	3.9	783	29	CNS00AIS	AL055833 Drosophil
c 18		52	3.9	797	13	BX437575	BX437575 BX437575
19		52	3.9	1101	29	CNS0181N	AL108773 Drosophil
c 20		51.8	3.9	623	28	AQ157491	AQ157491 nbxb0009H
c 21		51.8	3.9	830	29	CNS06P7Z	AL409029 T7 end of
c 22		51.8	3.9	1099	13	BX456575	BX456575 BX456575
c 23		51.6	3.9	946	13	BX416277	BX416277 BX416277
24		51.6	3.9	1201	13	BX461128	BX461128 BX461128
c 25		51.4	3.9	1080	28	AQ738730	AQ738730 HS_5382_B
c 26		51.4	3.9	1101	29	CNS0039R	AL063932 Drosophil
c 27		51.2	3.9	997	29	CNS005TE	AL060767 Drosophil
c 28		51.2	3.9	1204	29	CNS016E2	AL106628 Drosophil
29		51	3.9	673	29	CNS03XD5	AL264866 Tetraodon
c 30		50.8	3.8	1011	13	BX349844	BX349844 BX349844
c 31		50.6	3.8	929	13	BX408907	BX408907 BX408907
c 32		50.6	3.8	938	29	CNS006TJ	AL065906 Drosophil
c 33		50.4	3.8	626	28	AZ020346	AZ020346 RPCI-23-3
34		50.4	3.8	924	29	CNS01GHN	AL143148 Anopheles
c 35		50.4	3.8	1001	29	CNS0064G	AL062781 Drosophil
c 36		50.2	3.8	1101	29	CNS017KX	AL108171 Drosophil
37		50	3.8	1049	13	BX353130	BX353130 BX353130
38		50	3.8	1101	29	CNS0106X	AL098595 Drosophil
c 39		50	3.8	1201	13	BX335607	BX335607 BX335607
40		50	3.8	1201	13	BX385867	BX385867 BX385867
41		49.6	3.8	658	28	AZ851742	AZ851742 2M0154E07
42		49.6	3.8	735	29	AG095704	AG095704 Pan trogl
43		49.6	3.8	1201	13	BX460090	BX460090 BX460090
44		49.2	3.7	732	28	BZ146517	BZ146517 CH230-453
45		49.2	3.7	760	29	AG056350	AG056350 Pan trogl

ALIGNMENTS

RESULT 1
AW754659
LOCUS AW754659 526 bp mRNA linear EST 01-MAY-2000
DEFINITION PC04G10 Pine TriplEx pollen cone library Pinus taeda cDNA clone PC04G10, mRNA sequence.
ACCESSION AW754659
VERSION AW754659.1 GI:7676379
KEYWORDS EST.
SOURCE Pinus taeda (loblolly pine)
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 526)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh ,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda TriplEx2 Sequencing Primer.
FEATURES Location/Qualifiers
source 1. .526

/organism="Pinus taeda"
/mol_type="mRNA"
/db_xref="taxon:3352"
/clone="PC04G10"
/dev_stage="immature"
/lab_host="E. coli BM25.8"
/clone_lib="Pine TriplEx pollen cone library"
/note="Organ: pollen cone; Vector: Lambda TriplEx; Site_1: SfiI (A); Site_2: SfiI (B); Immature pollen cones were collected in the early spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda TriplEx vector. Plasmid subclones in pTriplEx were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 134 a 99 c 128 g 146 t 19 others
ORIGIN

Query Match 11.7%; Score 154; DB 9; Length 526;
Best Local Similarity 97.6%; Pred. No. 4.5e-15;
Matches 165; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy        936 GGAAAGGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATC 995
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db          2 GGAAAGGAAGAGGGANNNGGCATATGTATTGATCAACCTACCCGAAAAAACAATCTGATC 61

Qy        996 AGCCCTGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTC 1055
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGCCCTGCT-AATCTTGCTTATAAATCTCTTATCCACTGTTCAATCATTCAGGTTTCTTC 120

Qy       1056 CCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAG 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCACTTTCAAGCAAAGGCGCCCGGATTGGCCGTGTTCTTAGATTTTCAG 169
```

RESULT 2
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 4.7%; Score 61.4; DB 29; Length 1101;
Best Local Similarity 17.9%; Pred. No. 2.3;
Matches 140; Conservative 295; Mismatches 339; Indels 9; Gaps 1;

```
Qy         17 ATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCATATACATAACAAGCGGT 76
              :| : | ::|::| ::    :::::||  |:   : :  ::  : ::  : : : : |::
Db       1013 WTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRKDGDDDKKGGKK 954

Qy         77 GATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTA 136
              :  |:| :    ::  : :::: ::||::: |::::| |:  :  : : |:::   |::
Db        953 KKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTKD 894

Qy        137 GTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTT 196
              ::::::::: :: :| : :  :    ::::: : : :: |::| |::::|::: :|: :
Db        893 DDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDAWAW 834

Qy        197 GACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTG 256
              :: : :|   :  :: :: ::| ::    ::  :::  ::::| :: ::   :|  | |
Db        833 KWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWTTTT 774

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              :: :   :  :|:|::||:: :::::::::: |  | ::::: : : :::::|:::: :
Db        773 TRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDAD 714

Qy        317 AGTGAAAACG---------GTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCG 367
              |:: |::: :          |:|::  ::::: : | : : :| : :| ::|    :|:
Db        713 ADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDW 654

Qy        368 CAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAG 427
              :  :::| ::::     :| :: :|:: :: |  : : : :  ::: |:|  : ::|::
Db        653 DKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADR 594

Qy        428 AAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACG 487
              ::|:|:| :|:::: :  |: : ::||:     |  :|   | :     | :|:||  |
Db        593 DWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTTT 534

Qy        488 AATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGG 547
              : | : |     | |  :   | : :  ::|     :| | :   | ||:  | | | |
Db        533 WTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAAWAATTTTTTTTT 474

Qy        548 GGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGATTGCAGT 607
              |   |  : |:     |    |: |    | : :|     |  |||  :||
Db        473 TTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTTWTTTWTAA 414

Qy        608 GTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATCCTCTC 667
              | ::   : || | ||  |    ||  :| :|||    ::     :| :|| :::
Db        413 AAAAAAWWWWWTWTATATKCCCCCCYCCCCCCYCCYCCCCTCCYYCCTCCYTCYCCCYYY 354

Qy        668 CATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTCCACTG 727
              : :: :: :| ||    ::: : |||  :   : : :::: |:|     |::::|| :|
Db        353 YYYYYYYYYTYCCCCTYYYYYYCCCTTTYYYYYYYCYYYYCCYTCNCNCCYYYYCCTYTY 294

Qy        728 CCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTTTCCCC 787
              |::   : | |      | :|: :|  |:: :   :   : :| |||| || : :|::
Db        293 CYYYTTYTCCNCNCCYCCCCYCYYYCCYCYYCYCCCYYTYYCYTTTTTTYTTTYCYCYYT 234

Qy        788 TCC 790
              :|:
Db        233 YCY 231
```

RESULT 3
CNS00418
LOCUS CNS00418 987 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066537
VERSION AL066537.1 GI:4942778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 987)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .987
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR09C16"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 238 a 162 c 17 g 177 t 393 others
ORIGIN

Query Match 4.3%; Score 57.2; DB 29; Length 987;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 180; Conservative 94; Mismatches 386; Indels 5; Gaps 1;

```
Qy        258 GAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAA 317
Db         52 GAAAAAVAAAAAAANNDAWCGANANNNNANAAANAAAAAANNNAATATATAANAAAAAAADA 111

Qy        318 GTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGC 377
Db        112 ATAANAAAAAANAAAAAAAAATAAACAAAYAAAASHGNNNNNNGAGGGAANANNAANTNN 171

Qy        378 AGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGA 437
Db        172 NNNAANNCTTNNANNGAANNAAANNNANAATACAANAAAAAAAAANAAAAAAAAAANAAAA 231

Qy        438 AGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAA 497
Db        232 AAAAAANAAAAAANTNAAGAANANAANATNNATATAAATAGNGAAAANAAAAAAAAAAAA 291

Qy        498 ATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCGGCATT 557
Db        292 AAAAAAAAAAAAAAAAANAAAANAANAANNNANAANNAAAANTNAATNNNCNCCSTAACKN 351

Qy        558 CACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGATTGCAGTGTGGGAGTTG 617
Db        352 NANNNNNNAAANNNANCCANAGANNANNNAANNNNNNNAAAANNNNAAAAGANANNNAAA 411

Qy        618 CAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATCCTCTCCATTACTATT 677
Db        412 NANNAANNAANGANNANGNNATCYTCCYCCYTCYCCYTYYYCTTYYCTCYTCCYTCCCYY 471
```

```
Qy        678 ACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTCCACTGCCTGCACACG 737
Db        472 CYYYYYCYCTYTCTCTTCTTYTTYYTTCTYYCCYYYYTYTYYCTYYCCYYYYYYCYCYCC 531

Qy        738 TCTTAGTCAATCCATCTGCCTTCAAATAGGCATT-----ATTTTGTTCTTTCCCCTCCGA 792
Db        532 YCYYCYCCCCCCCYYCCYYCYCCYCCYCTCCCTYTCCYCYTTTYTTTYYCTYYCYYTYYY 591

Qy        793 CTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTG 852
Db        592 YTCYYTCTCTYYTYYYCYYYCYCYYTCYCTCTYYCYTYYYYYTCTYCCYCYYCYCCTCTC 651

Qy        853 GATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATC 912
Db        652 YYYCCYCYTCYYYYYCTCYYYYTCYCCYCYCCYCCYCYYCCYCTCTCCCTCYCYCYTCYY 711

Qy        913 TCTCC 917
Db        712 TYTYC 716
```

RESULT 4
BX461310
LOCUS BX461310 1201 bp mRNA linear EST 22-MAY-2003
DEFINITION BX461310 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF025YA16 5-PRIME, mRNA sequence.
ACCESSION BX461310
VERSION BX461310.1 GI:31031291
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF025BA08QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF025YA16"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 472 a 108 c 138 g 167 t 316 others
ORIGIN

Query Match 4.3%; Score 56.2; DB 13; Length 1201;
Best Local Similarity 21.6%; Pred. No. 15;
Matches 81; Conservative 139; Mismatches 155; Indels 0; Gaps 0;

```
Qy        141 GGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACT 200
Db        782 RRGRGGGGGGGGWWWDDDKKKKKKKKKKKKTKKKKKKDAAAAAAAAAAAAKDKDKDARADD 841

Qy        201 CTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAA 260
```

```
Db        842 GGGGKKKKKKDDAAAAADDKKKKKKKKKKKKKKKKKRAAAAWTTKKKKKKKKKKDKKKKK 901

Qy        261 GCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTG 320

Db        902 KKKKKKKDAAAAAAAAAAAAAAADDKDDAAAAAAAAAAAAAAKKKKKKKKKKKKKDAAAAA 961

Qy        321 AAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGA 380

Db        962 AAADKKKKKKKKKKDKKKDAAAAAAAAAAAAAAAAAAAAAADDKKKKKKKKKKKKKK 1021

Qy        381 TTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGA 440

Db       1022 KKKKKKKKKDAADKKKKKKAAAAAAAAAAAAAAKKDAAADDKKKKKDAAAAADKKKKKK 1081

Qy        441 GTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATA 500

Db       1082 AAAAAAAADKDDDAAAAAKKKKKKKKKKDAAAAAADDDKDKKKKKKKDDKKKAAAAAAAA 1141

Qy        501 TAATGCAAGAGTGCA 515

Db       1142 AAAAAAKAAAKKKAA 1156
```

RESULT 5
CNS003B2
LOCUS CNS003B2 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR08A08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL064080
VERSION AL064080.1 GI:4941836
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08A08"
/clone_lib="RPCI-98"
/note="end : T7"
BASE COUNT 399 a 197 c 187 g 196 t 122 others
ORIGIN

Query Match 4.2%; Score 55.2; DB 29; Length 1101;
Best Local Similarity 45.5%; Pred. No. 22;
Matches 117; Conservative 17; Mismatches 123; Indels 0; Gaps 0;

```
Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316

Db        844 AAAAAWGAAAAAATTRTATKRAAAAAAAWAAAAAAAAAARAAAGAAAGAGGRAAAAATGAA 903

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376

Db        904 AAAAAAAAAAATAAAAAAAAAAAAAAWTAAAAAAAAAAAAAAAAAWAWNGAAAAAGAGAAA 963

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436

Db        964 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAAANGAAAATGAGT 1023

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496

Db       1024 GAKTGKGAAKAAWAAATGAARATGANGTAAGARAGAAAAAAAAAARAGARAGAATGGAARA 1083

Qy        497 AATATAATGCAAGAGTG 513

Db       1084 AAAAAAAAAAAAGWGKG 1100
```

RESULT 6
CNS00DG1/c
LOCUS CNS00DG1 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR27J03 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069971
VERSION AL069971.1 GI:4947993
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27J03"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 227 a 174 c 33 g 361 t 306 others
ORIGIN

Query Match 4.2%; Score 55.2; DB 29; Length 1101;
Best Local Similarity 34.3%; Pred. No. 22;
Matches 114; Conservative 62; Mismatches 156; Indels 0; Gaps 0;

```
Qy        173 GAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGG 232
```

```
Db        952 KAGGWAGAAAARGARAAAAAARARRAGGAKTGGAAAGGTGGAARGRGATTGKTAAAGRGR 893

Qy        233 GCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAA 292

Db        892 GGATRAAAWVTTTWTSWAAAAAVDTSSTSAAAAASAAAAVARAAAAAAAAAAAAAAAAAA 833

Qy        293 AAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAG 352

Db        832 WAAAAAAATWAAAAAAASTSKTAAAAAASSSAVAASAWWWATGSSRATAAAWRARSARGRA 773

Qy        353 TACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGA 412

Db        772 AARTWTAAAAAAAAGAARAAAAAASSVASASWAWRAVSSASAAAWSAAWAAARVAAAATT 713

Qy        413 GATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGC 472

Db        712 GATAARGGTGGGAGAAAARTTWWAATTAAWGWWRRAATGRTAWRARARAATATGAARSAG 653

Qy        473 ACAAAATGGAGGACGAATAAATGAAATATAAT 504

Db        652 ARTWAATWRAAAATGWAAAATTTARTTTGAAT 621


RESULT 7
AL559034/c
LOCUS       AL559034                1041 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL559034 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ010YL15 5-PRIME, mRNA sequence.
ACCESSION   AL559034
VERSION     AL559034.2  GI:31283167
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12904134.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            8934.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DJ010CF08QP1&cluster=8934.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DJ010CF08QP1.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ010YL15"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      153 a    190 c    201 g    335 t    162 others
ORIGIN

  Query Match             4.2%;  Score 54.8;  DB 9;  Length 1041;
  Best Local Similarity   33.7%;  Pred. No. 26;
  Matches   86;  Conservative   29;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        254 GTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTT 313

Db        460 GAGTRWAAAAARGRCCTGWRAAAAAAAAAAARAATATAAGAAATGAAAAAARWAAAAGGAA 401

Qy        314 GGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCA 373

Db        400 AAAARAAAAWAAAWGRARRARRAAAAWAAAAAAARAAAAWAAWAAAAARAWARAAAAAAA 341

Qy        374 AGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAA 433

Db        340 NNAAAAANNNNNNANNANNANNNNNNNNNANNAAAAAAANNNNANNNNGGAANNNAANNA 281

Qy        434 AGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAA 493

Db        280 ANNAANNANNNNNNANNNNNNNNANNANNTGAGGGCGRRARMCCGRGCARAGRAGCCAR 221

Qy        494 TGAAATATAATGCAA 508

Db        220 RCARATCCAACGARM 206


RESULT 8
CNS006N9
LOCUS       CNS006N9                964 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065781
VERSION     AL065781.1  GI:4944661
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14J19"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      409 a    111 c     71 g    130 t    243 others
ORIGIN

  Query Match             4.1%;  Score 54;  DB 29;  Length 964;
  Best Local Similarity   34.0%;  Pred. No. 36;
  Matches   86;  Conservative   64;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        264 AACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAA 323
```

```
Db        664 AAAAATTGTAGTWAWTAKTAAAAAATATAAAAAADAAADAADAAAKKTKWDKAADAAAAA 723

Qy        324 ACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTC 383

Db        724 AAGWWAKKDAAAAAAAAAAAAAAAKAKGTATADADAAAAADTAKAKWDWADAKDDDKKAGA 783

Qy        384 ACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTG 443

Db        784 AKADKWARKAAARAAAAWCKKDARKAKAWAAKARWAKAGTGWAAKARAGKATAGAGAGAK 843

Qy        444 GAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db        844 AAKAKDKDKAGKDAKADAAKAKADAKAKADKADDAKAGADKAGAKAKKKGKANAGDAKAK 903

Qy        504 TGCAAGAGTGCAT 516

Db        904 AKAAAAAAARWAK 916
```

RESULT 9
CNS0174K
LOCUS CNS0174K 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN17P20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL107582
VERSION AL107582.1 GI:5627886
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN17P20"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 365 a 90 c 235 g 206 t 205 others
ORIGIN

Query Match 4.1%; Score 54; DB 29; Length 1101;
Best Local Similarity 42.5%; Pred. No. 35;
Matches 147; Conservative 22; Mismatches 177; Indels 0; Gaps 0;

```
Qy        253 TGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGT 312

Db        216 TATAATAGRATAATAATAATTAGGGWTAAAATAAAAAAAWATAAAGAAATAAAAAARGAA 275

Qy        313 TGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACC 372

Db        276 AATTAAAAAAAAAAARGWAGAAAAATATTAAAAAAAAGTARWTATTWAAAAATAATGAAAT 335

Qy        373 AAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATA 432
```

```
Db        336 AAWTTATATWTAAAAAARGAAAGATGRAWAGRAAAATRAAAAWAAAAATRATAAAAAATA 395

Qy        433 AAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAA 492

Db        396 AWTWAAAAAAAWWATTAATATAAATAAATAAAAAAAAAGTTAAATATAAAATTAAAAAAT 455

Qy        493 ATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCG 552

Db        456 AAGTATATAAAAAAAATWGAGTTTGATTSGAATAAAGTGATAGATGATTKAGAGGCGGAG 515

Qy        553 GCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTT 598

Db        516 GCAMTCAMGAASGTCGCKAAAWAGCGGTCACCGGTATCSATTCTYT 561
```

RESULT 10
CNS017KE/c
LOCUS CNS017KE 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN37F11 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108152
VERSION AL108152.1 GI:5628456
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F11"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
BASE COUNT 302 a 83 c 32 g 379 t 305 others
ORIGIN

Query Match 4.1%; Score 54; DB 29; Length 1101;
Best Local Similarity 35.1%; Pred. No. 35;
Matches 99; Conservative 57; Mismatches 126; Indels 0; Gaps 0;

```
Qy        264 AACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAA 323

Db        711 MAGDAAKKGAAAAAAAAAAAHDADAAAAAAAAAAAGGWAAAAKKAKKKKKMAGRAKAKACA 652

Qy        324 ACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTC 383

Db        651 AAKKAAGKTCAGACACKAKAAAAAAGDCKKMGGKDAKAAARAKMKAAAKKAKKAKKKAKK 592

Qy        384 ACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTG 443

Db        591 KKKAKKAAAKKDWCRAAKKWAAAAKDKKAGAGAADKKGKAAAAGAAKAKAAATATAWSGA 532

Qy        444 GAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503
```

```
Db        531 TTGKWCAGKATAAGAWKAAKAAAAAKARGKAAKWATCKAKAAKKAKAAAAKATGAATATA 472

Qy        504 TGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGT 545
                  |  ||   | |   |||| |||          ||||| |
Db        471 ATTTAATGTTTTTGTTATTATTTTTTTTTATTTATTATATAT 430


RESULT 11
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      195 a    108 c    131 g    161 t    506 others
ORIGIN

  Query Match             4.0%;  Score 53.4;  DB 29;  Length 1101;
  Best Local Similarity   18.2%;  Pred. No. 43;
  Matches   82;  Conservative  187;  Mismatches  179;  Indels    3;  Gaps    1;

Qy         96 TGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTG 155
              :: : : : ||:    : ::::  : :  |  :  : ::: ::: :|:: :|: ::  ||
Db        632 KRGWGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWGWKDKWTDKKRDGKRGDGDRKDKGTG 691

Qy        156 GCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATT 215
              |  ::: :|::|:|::|:| : |||:|   | ||: ||: |:  | :  |    :   ||
Db        692 GDGDSKAKGDRAWAWRAKATRAAAADATAAATAAKATAWATKTTTKKTTTTTTTKTTTTT 751

Qy        216 GGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAA 275
                ||  |||    ::: :::   :| ::|: :: :::::||:::  :   : ::: |
Db        752 TTTTTTKKTTTARKADRDDDDKAKRDGDKDGDWDDADKDWKAGRDRDDDTRKWGWKDRAG 811

Qy        276 GAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAG 335
              :    : ::  |:::|::   |:| :|:|  |:||: :   :::: :  || |:|: ||
Db        812 KKKKAKKKDHKTAWWWADDTGTAWATWAKTATAKTGKAKATKDBKRGRGGGGRGDGRGAG 871

Qy        336 ---AAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392
                 :|: ::: : ::: : ::  |::|   :|| | :   : :: :|: : : :  : :
Db        872 RRKRARGRKRGRRRKDKRKRKKRTKKAGDGDTGAGTDKAGDKKADGDGWGWKRKWDSKT 931

Qy        393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452
              ::|   |   |  :  : ::|: :|  | :: : | | : | |::::: : :|::   ::

Db        932 WKTTMTGTDDTTTWKGRNRRGRGRAGAGRRRGRKGKAGKGAGGKWWKWAWADABWTWTWD 991

Qy        453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGT 512
              :::: ::  ::|  : | |       |:||::  : :: :::| :|: ::::   |:|::
Db        992 WKDWKWKWTDWGWGKGGWGDGRKGGAKDGAKRNRKNRDGMRYGNRAWKWMDDAATAKARW 1051

Qy        513 GCATTTCCCTATTATTTCCAGAAATGTATAT 543
                 ::::  :::: : ::  ::| : ::: :|
Db       1052 ADDDWDADYDDDADADYWKWKATWMDBWCDT 1082


RESULT 12
CNS005TE
LOCUS       CNS005TE                997 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       89 a     99 c     13 g    258 t    538 others
ORIGIN

  Query Match             4.0%;  Score 52.8;  DB 29;  Length 997;
  Best Local Similarity   20.6%;  Pred. No. 56;
  Matches   61;  Conservative  111;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        635 TCCATTCATCCCAAAACCTAAATTTATCCTCTCCATTACTATTACCTACACCTATACCTA 694
              :|: | : ::|:     | :     :|  :::: :: :: ::  : :::     |: | :::
Db        595 YCYYTCYTYYCYYYCYCTYYYYCYTYCYYYYCYYCYYYYYTCYTYYYTTCTCYYTYYYYY 654

Qy        695 GTAAATATGTCCTGCCTTGTAACTCCTCCACTGCCTGCACACGTCTTAGTCAATCCATCT 754
                | : : ::::  ||  :  : :|::: || :|: | : : |||    |  |:: |::
Db        655 TYYAYKYYCYYYYYTCTCYYYCYCYCYYYTCTCYCYYCYYYYTTCTCTYTTYYTYYYTY 714

Qy        755 GCCTTCAAATAGGCATTATTTTGTTCTTTCCCCTCCGACTGAAAGGCTATCGACCGACCG 814
              :  ||    :   : :| :||| |::|::::  |:   :|       | ::  ::  |:
Db        715 TYTCTCYYYYYYYYYYTYYTTTCTYYTYYYYTTTYTCTYTYTYTYYTTYYYYCYYYCCYC 774
```

```
Qy        815 ACCGCTCATCTTCTTCTTCTGCGCAATTTTTCTGCTGGATCATCATCATTACCATCATC 874
                  :: |:: :|:||:|::::: :   |:|:||:|  :   |: || :: |: :: : :|
Db        775 YYYCCYYYYCYTCYTYYYYYCTYYYYTYTYTTYTYYCYTYTYTTCTYYYTYTYYYCYCYC 834

Qy        875 GCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATT 930
                  :| :|| |  : :| :| :: : :  | : : |:|:|::: :|  :  |: ::|
Db        835 YYCTYCCTCYTYYYCTYCYYYYYCYYYYTCYTYTTMTYTYYYTYTYTYTYTYTHYT 890

RESULT 13
BX122696
LOCUS       BX122696                 634 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-58A23, genomic survey sequence.
ACCESSION   BX122696
VERSION     BX122696.1  GI:27953626
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 58A23. 58A23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1..634
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-58A23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
BASE COUNT      363 a     75 c    151 g     45 t
ORIGIN

  Query Match             4.0%;  Score 52.6;  DB 29;  Length 634;
  Best Local Similarity   50.2%;  Pred. No. 70;
  Matches  130;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              | |||  | | |   | ||   | ||  |||  || |||||||| ||       ||
Db        298 AAAAGGAATCTAAGGAAAAATCAAATGGAAACAAAGAAATAAAATAAAAGTTTAACTGAA 357

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376
              | |||| ||  |  | ||||  | | | |||||         |    |||     |   ||
Db        358 AATGAAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA 417

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436
              ||||    |||    ||          ||||        |||   |       |||||| ||||
Db        418 CAGACAGACAGACAGACAGACAGACCGACCGACCGACCGACAAACAGAAAGAAAGAAAGA 477

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496
              ||||  |  |     |      |||  ||    |||  |  |  |||    || | |||  || ||
Db        478 AAGAAAGAAAGAAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 537

Qy        497 AATATAATGCAAGAGTGCA 515
              || | || | |||| | |
Db        538 AAGAAAAAGAAAGAAAGAA 556

RESULT 14
CNS0182P
LOCUS       CNS0182P                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108811
VERSION     AL108811.1  GI:5629115
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37D10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      274 a    268 c    128 g     73 t    358 others
ORIGIN

  Query Match             4.0%;  Score 52.6;  DB 29;  Length 1101;
  Best Local Similarity   21.7%;  Pred. No. 58;
  Matches   75;  Conservative  132;  Mismatches  138;  Indels    0;  Gaps    0;

Qy        254 GTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTT 313
               :|| |  |   :  ||  :|:::|: : : | ||:| |||:| ||||||| ||
Db        654 KTGKGKAADGDGGAAAADAWRRGRGRGRRARAARAAAAADAKRAAAAAAAAAAAAAAAAA 713

Qy        314 GGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCA 373
               :| |:: :   :|::||:  : ||:|:   :| | | :  : |:|  ::     :  :
Db        714 ADAGGKRKWKKGDGKDGGKATKTAAWAKGRKDWGTATAWWTWDATWADTWKAATDTDAKR 773

Qy        374 AGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAA 433
              |  :: : :    | : |:    ::|   | : ||:: :| :: ::: :|: ||
Db        774 AAAGRRKRDARKTARDGGRRARTRRRAWAAGGKRARAGARRARAARRAADDRDWDAWAAA 833

Qy        434 AGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAA 493
              | ||  |     :: :::: ::::|:::::   |  :  | |:|: :::::: ::: ::|
Db        834 AAAAAAAAAAWTTWRDRWWDWDDWDTRWDDDTTAAWWDDARARARRRRRRRRRRRARRA 893

Qy        494 TGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCGG 553
              ::||::|: :: :|:| |: :||:  |::::::: ::::: :|:   : ::: ::
Db        894 ARRAADDTDTKDRWADATTDKDTTKTWTTDDDDWDKAKRDRWWAAKADGAWKWRDRARDD 953

Qy        554 CATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTT 598
               |   ::| : ::| |  :   :::::::: | : ::: ::|
Db        954 WAATAKDDDGWKDKWGGRGKRGGKDKKRWDKGKTGKKDDDDWDKT 998

RESULT 15
BX407949/c
LOCUS       BX407949                1114 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX407949 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM001YI02 3-PRIME, mRNA sequence.
ACCESSION   BX407949
VERSION     BX407949.1  GI:30768711
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1114)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6978.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AS014ZE09NP1&cluster=6978.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AS014ZE09NP1.
FEATURES              Location/Qualifiers
     source           1. .1114
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DM001YI02"
                      /tissue_type="FETAL LIVER"
                      /dev_stage="fetal"
                      /clone_lib="Homo sapiens FETAL LIVER"
                      /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                      was primed with a NotI-oligo(dT) primer. Five prime end
                      enriched, double-strand cDNA was digested with Not I and
                      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                      vector. Library was not normalized."
BASE COUNT      139 a    161 c     51 g    670 t     93 others
ORIGIN

  Query Match             4.0%;  Score 52.6;  DB 13;  Length 1114;
  Best Local Similarity   44.2%;  Pred. No. 58;
  Matches  144;  Conservative   17;  Mismatches  164;  Indels    1;  Gaps    1;

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              |:||   ||:||   : |||||: | | |||| ||||| |||||||    ||   ||| |
Db        399 ARAAAAAAAMAAMCAMAAAGAARAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAATTGAA 340

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376
              || |:|  | |||    |    | |    ::   |:        ||   | ||  :|
Db        339 AGGGRAGGCNGTGTAAGANTTGTTTNTGGRSGGNASAGTGNNTTAAAAAAGAAAAARAAA 280

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436
               | |   | ||   ||    ||       ||      | |  ||||||||   || ||| |
Db        279 AAAAAAAAAAAAAAGAAAAGTGTNTTGAGATAGNNNGGTNNTGGTGGTATTTAAAAAAAG 220

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496
              :| |    :        |:||| :|| || | |   | |||| | | |   || ||| ||
Db        219 MAAAAAAAMAAAAGAAAGRAAAARAAAGG-GGGAAGAAAAAAGGAAAGGAAAAAAAAAGA 161

Qy        497 AATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCGGCAT 556
              |: | |||   | || | |      |  |     | | |||   | |   | |
Db        160 ARGAAAATAAGAAAGGGAAAAAAAAAAAAAAAAAAAAAAANATGGGGAAGGTAGTTTTTTCA 101

Qy        557 TCACATGGGCGTCGCATTCAGGGGGT 582
              ||  | : |  ||  || |||||::|
Db        100 TCTAAARTGTCTCTGATGCAGGGKRT 75


Search completed: January 22, 2004, 00:54:15
Job time : 2640.89 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 16:52:57 ; Search time 89.2891 Seconds
(without alignments)
6525.161 Million cell updates/sec

Title: US-09-936-869-1
Perfect score: 1320
Sequence: 1 aaactcgacagcaaatatga.....tgttgctgttgcgcccattg 1320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	73.4	5.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	61	4.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	3	43.6	3.3	176373	3	US-09-128-155-17	Sequence 17, Appl
	4	42	3.2	152331	3	US-09-128-155-16	Sequence 16, Appl
c	5	41	3.1	2255	2	US-08-741-134-1	Sequence 1, Appli
c	6	40	3.0	87350	3	US-08-781-891-79	Sequence 79, Appl
c	7	40	3.0	87350	4	US-09-618-166-79	Sequence 79, Appl
c	8	40	3.0	87543	4	US-09-791-211-3	Sequence 3, Appli
c	9	39.4	3.0	435	4	US-09-397-787-171	Sequence 171, App
c	10	39.4	3.0	1024	4	US-09-328-475C-64	Sequence 64, Appl
	11	39.2	3.0	1753	3	US-09-058-260-21	Sequence 21, Appl
	12	39.2	3.0	1756	3	US-09-058-260-31	Sequence 31, Appl
	13	39.2	3.0	1776	3	US-09-058-260-23	Sequence 23, Appl
	14	39.2	3.0	1952	3	US-09-058-260-5	Sequence 5, Appli
	15	39.2	3.0	4090	2	US-08-781-802-5	Sequence 5, Appli
	16	39.2	3.0	4090	3	US-08-694-078-5	Sequence 5, Appli
c	17	38.8	2.9	3564	2	US-07-594-921C-9	Sequence 9, Appli
c	18	38.8	2.9	3564	4	US-09-197-948-9	Sequence 9, Appli
c	19	38.6	2.9	202001	4	US-09-734-674-3	Sequence 3, Appli
	20	38.2	2.9	289	3	US-09-007-005-17	Sequence 17, Appl
	21	38.2	2.9	289	3	US-09-244-796-17	Sequence 17, Appl
	22	37.6	2.8	1896	3	US-09-058-260-3	Sequence 3, Appli
	23	37.6	2.8	6263	2	US-08-781-802-3	Sequence 3, Appli
	24	37.6	2.8	6263	3	US-08-694-078-3	Sequence 3, Appli
	25	37.4	2.8	475	4	US-09-702-705-36	Sequence 36, Appl
	26	37.4	2.8	475	4	US-09-736-457-36	Sequence 36, Appl
c	27	37.4	2.8	4235	4	US-09-702-705-317	Sequence 317, App
c	28	37.4	2.8	4235	4	US-09-736-457-317	Sequence 317, App
c	29	37.4	2.8	72604	4	US-09-268-992-7	Sequence 7, Appli
c	30	37.4	2.8	72604	4	US-09-657-474-7	Sequence 7, Appli
c	31	36.8	2.8	289	3	US-09-007-005-17	Sequence 17, Appl
c	32	36.8	2.8	289	3	US-09-244-796-17	Sequence 17, Appl
	33	36.8	2.8	246240	2	US-08-724-394A-20	Sequence 20, Appl
	34	36.8	2.8	246240	2	US-08-724-394A-21	Sequence 21, Appl
	35	36.8	2.8	246240	2	US-08-724-394A-22	Sequence 22, Appl
	36	36.6	2.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	37	36.4	2.8	658	3	US-08-998-416-595	Sequence 595, App
c	38	36.4	2.8	1164	4	US-09-538-871-3	Sequence 3, Appli
c	39	36.4	2.8	1232	4	US-09-538-871-1	Sequence 1, Appli
c	40	36.2	2.7	2013	4	US-09-134-001C-103	Sequence 103, App
	41	36.2	2.7	176373	3	US-09-128-155-17	Sequence 17, Appl
c	42	36	2.7	2518	3	US-09-433-699-3	Sequence 3, Appli
c	43	36	2.7	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c	44	35.8	2.7	1141	2	US-08-323-449B-1	Sequence 1, Appli
c	45	35.8	2.7	1141	2	US-08-485-981-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match 5.6%;  Score 73.4;  DB 1;  Length 7218;
```

```
  Best Local Similarity   6.2%;  Pred. No. 9.9e-11;
  Matches   26;  Conservative  236;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          93 CACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGC 152
               ||||||| ||    |   |   |    || || |  | |||||| ::  ::
Db        1487 CACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRR 1428

Qy         153 TTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTAC 212
                 ::     ::: :: : :::::: ::::::: :::::: :   ::    ::  ::: :
Db        1427 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1368

Qy         213 ATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTA 272
               :  :::  :     :  ::::   ::: ::   :::   : : ::::::  :: ::   :
Db        1367 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1308

Qy         273 TAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGG 332
                ::::::  : : :::: :::::: :::::::   ::: :  ::::: ::::: :: ::::
Db        1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248

Qy         333 CAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392
                ::::: : : : :::::::  :    :: :: : : ::  :::: :::   : ::   ::
Db        1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188

Qy         393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452
                  :::    ::: : :::: :: :: :: :::::: :::::::::: :::: : :
Db        1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128

Qy         453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAG 511
                ::::: ::: :::::: : : :::: :::::: ::: ::: :::: : :: : :::::
Db        1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069


RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.6%;  Score 61;  DB 1;  Length 7218;
  Best Local Similarity   2.2%;  Pred. No. 3.3e-07;
  Matches    7;  Conservative  200;  Mismatches  110;  Indels    0;  Gaps    0;

Qy         603 GCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATC 662
               | ||  || || :: :  : : : ::  :   ::: ::: ::::    :::   ::: ::
Db        1055 GGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114

Qy         663 CTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTC 722
               :::::: :: :: :: ::: : ::: : ::: :  :  : : :::: :::: :  ::::::
Db        1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

Qy         723 CACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTT 782
               : :: ::: : : : :::  ::  ::: ::: :::::   :  : :: :::: :::::
Db        1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

Qy         783 TCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATT 842
               ::::::::  ::      :: ::  ::  ::  :: ::: :::::::::::: : :  ::
Db        1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294

Qy         843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
               :::::: ::   :: :: :: :: :: :: :: :: :: ::::: :: :: :: :  :
Db        1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354

Qy         903 TATCTCTATCTCTCCCT 919
               : ::::: ::::::::
Db        1355 YYYYYYYYYYYYYYYYY 1371


RESULT 3
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

  Query Match             3.3%;  Score 43.6;  DB 3;  Length 176373;
  Best Local Similarity   48.8%;  Pred. No. 0.14;
  Matches  118;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
               |||| | | |||| ||| ||  ||| ||   |||           | |||  | |
Db      122513 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 122454

Qy         334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
```

```
Db     122453 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 122394

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453

Db     122393 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 122334

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513

Db     122333 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 122274

Qy        514 CA 515

Db     122273 AA 122272


RESULT 4
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             3.2%;  Score 42;  DB 3;  Length 152331;
  Best Local Similarity   48.3%;  Pred. No. 0.37;
  Matches  117;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333

Db     146880 AGGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 146939

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393

Db     146940 GAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 146999

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453

Db     147000 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 147059

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513

Db     147060 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGAAAAAGAAAGAAAGAAAG 147119

Qy        514 CA 515

Db     147120 AA 147121


RESULT 5
US-08-741-134-1/c
; Sequence 1, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,163
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2255 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  160..1398
US-08-741-134-1

  Query Match             3.1%;  Score 41;  DB 2;  Length 2255;
  Best Local Similarity   63.9%;  Pred. No. 0.09;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy        820 TCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATCATCGCCAT 879

Db        826 TCTTCTTCTTCTTGGGTGCTTCCTCCTCTTCTTCATCGTCCTCATCATCATCTTCTTCAT 767

Qy        880 CCCCACCATCATCATCATGATGGTATCTCTATCTCTC 916

Db        766 CCTCTTCGTCTTCATCATCATCCTCTTCATCACTGTC 730


RESULT 6
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  87350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-79

  Query Match             3.0%;  Score 40;  DB 3;  Length 87350;
  Best Local Similarity   50.8%;  Pred. No. 1;
  Matches  120;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328

Db      49802 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49743

Qy        329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387

Db      49742 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49683

Qy        388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447

Db      49682 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49623

Qy        448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db      49622 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49567


RESULT 7
US-09-618-166-79/c
; Sequence 79, Application US/09618166
; Patent No. 6583112
;   GENERAL INFORMATION:
;        APPLICANT: Fu, Ying-Hui
;                   Yu, Chang-En
;                   Oshima, Junko
;                   Mulligan, John T.
;                   Schellenberg, Gerald D.
;        TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                            WERNER'S SYNDROME
;        NUMBER OF SEQUENCES: 209
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/618,166
;             FILING DATE: 17-Jul-2000
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 240052.419C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 79:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 87350 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-618-166-79

  Query Match             3.0%;  Score 40;  DB 4;  Length 87350;
  Best Local Similarity   50.8%;  Pred. No. 1;
  Matches  120;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328

Db      49802 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49743

Qy        329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387

Db      49742 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49683

Qy        388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447

Db      49682 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49623

Qy        448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db      49622 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49567


RESULT 8
US-09-791-211-3/c
; Sequence 3, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 3
;   LENGTH: 87543
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 7421
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 7427
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 11609
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 12605
;   OTHER INFORMATION: unknown
```

; NAME/KEY: unsure
; LOCATION: 12742
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29370
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29422
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29979
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29980
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29981
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30136
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30140
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31205
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31206
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31592
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33095
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33160
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34066
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34072
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 36816
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 39020
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42164
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42459
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46808
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46823
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46826
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 47291
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52786
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52787
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 53384
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 54684
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59215
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59235
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59242
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 63290
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 66614
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68660
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68697
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68718
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68733
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68739
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 69785
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79134
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79198
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 86336
; OTHER INFORMATION: unknown
; OTHER INFORMATION:

US-09-791-211-3

Query Match 3.0%; Score 40; DB 4; Length 87543;
Best Local Similarity 50.8%; Pred. No. 1;
Matches 120; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy      269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328
            ||  ||| ||  | |||     | | | | |||||   || | ||  |  | ||||  ||
Db    49995 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49936

Qy      329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387
              |  |||  |  | | ||| | |   ||| | ||| | | ||  || |  |  ||||
Db    49935 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49876

Qy      388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447
              |   ||     ||      | | | || || |  |  || || | || ||  | || ||
Db    49875 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49816

Qy      448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503
              | |||| ||| | || | | || || || |  | |||   | ||  ||
Db    49815 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49760
```

```
RESULT 9
US-09-397-787-171/c
; Sequence 171, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 171
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-397-787-171

  Query Match             3.0%;  Score 39.4;  DB 4;  Length 435;
  Best Local Similarity   71.2%;  Pred. No. 0.11;
  Matches   52;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        825 TTCTTCTTCTGCGCAATTTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCA 884
              ||||||    ||||  ||| ||| |||   |||||||||  | |||||||  ||||  |
Db        372 TTCTTCACTGGCGCTTTTTCTTCAGCTTCCTCATCATCAAAATCATCATCATCATCATCT 313

Qy        885 CCATCATCATCAT 897
              ||||||||||| |
Db        312 TCATCATCATCCT 300


RESULT 10
US-09-328-475C-64/c
; Sequence 64, Application US/09328475C
; Patent No. 6476207
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Jimmy
;  APPLICANT:  Astel, Jon H.
;  APPLICANT:  Carroll III, Eddie
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Ford, Donna M.
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Steinmann, Kathleen E.
;  TITLE OF INVENTION: GENES AND GENE EXPRESSION PRODUCTS THAT
;  TITLE OF INVENTION:  ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER
;  FILE REFERENCE: 1532.002/200130.463
;  CURRENT APPLICATION NUMBER: US/09/328,475C
;  CURRENT FILING DATE:  1999-06-09
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
;   LENGTH: 1024
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1024)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-64

  Query Match             3.0%;  Score 39.4;  DB 4;  Length 1024;
  Best Local Similarity   71.2%;  Pred. No. 0.17;
  Matches   52;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        825 TTCTTCTTCTGCGCAATTTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCA 884
              ||||||    ||||  ||| ||| |||   |||||||||  | |||||||  ||||  |
Db        440 TTCTTCACTGGCGCTTTTTCTTCAGCTTCCTCATCATCAAAATCATCATCATCATCATCT 381

Qy        885 CCATCATCATCAT 897
              ||||||||||| |
Db        380 TCATCATCATCCT 368


RESULT 11
US-09-058-260-21
; Sequence 21, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 1753
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E013
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-21

  Query Match             3.0%;  Score 39.2;  DB 3;  Length 1753;
  Best Local Similarity   56.1%;  Pred. No. 0.26;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  ||| |  |||||||| |  |  |    ||  |   |||| ||| |       | ||
Db         24 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 83

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |  |||||  || || |||   |||||||||  ||    |||| |  || |||| |
Db         84 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 143

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        144 CGTTGGTGGAAA 155


RESULT 12
US-09-058-260-31
; Sequence 31, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
```

```
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 1756
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E027
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (130)..(1632)
US-09-058-260-31

  Query Match             3.0%;  Score 39.2;  DB 3;  Length 1756;
  Best Local Similarity   56.1%;  Pred. No. 0.26;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |      |  ||
Db         26 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 85

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |   | |||||  || || |||   ||||||||  ||    ||||  |  || |||| |
Db         86 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 145

Qy        325 CGGTGGGGCAGA 336
              || |||  | | |
Db        146 CGTTGGTGGAAA 157


RESULT 13
US-09-058-260-23
; Sequence 23, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 1776
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E015
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-23

  Query Match             3.0%;  Score 39.2;  DB 3;  Length 1776;
  Best Local Similarity   56.1%;  Pred. No. 0.26;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |      |  ||
Db         24 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 83

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |   | |||||  || || |||   ||||||||  ||    ||||  |  || |||| |
Db         84 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 143

Qy        325 CGGTGGGGCAGA 336
              || |||  | | |
Db        144 CGTTGGTGGAAA 155


RESULT 14
US-09-058-260-5
; Sequence 5, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1952
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E011
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (197)..(1699)
```

```
US-09-058-260-5

  Query Match             3.0%;  Score 39.2;  DB 3;  Length 1952;
  Best Local Similarity   56.1%;  Pred. No. 0.27;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |   ||  |   |||| ||| |     |  ||
Db         93 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 152

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |  | |||||  || ||  |||   ||||||||  ||    |||| |  || |||| |
Db        153 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 212

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        213 CGTTGGTGGAAA 224


RESULT 15
US-08-781-802-5
; Sequence 5, Application US/08781802
; Patent No. 5969121
;  GENERAL INFORMATION:
;    APPLICANT:  ALLEN, Larry
;    APPLICANT:  AIKENS, John
;    APPLICANT:  FONSTEIN, Michael
;    APPLICANT:  VONSTEIN, Veronika
;    APPLICANT:  DEMIRJIAN, David
;    APPLICANT:  CASADABAN, Malcolm
;    TITLE OF INVENTION:  Stable Biocatalysts for Ester Hydrolysis
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,802
;      FILING DATE:  10-JAN-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/694,078
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/019,580
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,704
;      FILING DATE:  11-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,995
;      FILING DATE:  01-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chao, Mark
;      REGISTRATION NUMBER:  37,293
;      REFERENCE/DOCKET NUMBER:  95,963-E
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4090 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  197..1699
;      OTHER INFORMATION:  /note= "E011 sequence of longest
;      OTHER INFORMATION:  open reading frame; other possible start codons ATG/met4;
;      OTHER INFORMATION:  TTG/leu7; GTG/val8; GTG/val15; GTG/val36; ATG/met62"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  197..1699
US-08-781-802-5

  Query Match             3.0%;  Score 39.2;  DB 2;  Length 4090;
  Best Local Similarity   56.1%;  Pred. No. 0.39;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |   ||  |   |||| ||| |     |  ||
Db         93 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 152

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |  | |||||  || ||  |||   ||||||||  ||    |||| |  || |||| |
Db        153 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 212

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        213 CGTTGGTGGAAA 224


Search completed: January 21, 2004, 22:19:40
Job time : 94.2891 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:11:45 ; Search time 444.739 Seconds
(without alignments)
10461.606 Million cell updates/sec

Title: US-09-936-869-1
Perfect score: 1320
Sequence: 1 aaactcgacagcaaatatga.....tgttgctgttgcgcccattg 1320

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	49.6	3.8	22684	13	US-10-017-161-725	Sequence 725, App
	2	49.6	3.8	29221	12	US-10-292-798-635	Sequence 635, App
c	3	45	3.4	519	10	US-09-878-574-4292	Sequence 4292, Ap
	4	43.8	3.3	625	13	US-10-027-632-44092	Sequence 44092, A
	5	43.8	3.3	625	13	US-10-027-632-44093	Sequence 44093, A
	6	43.8	3.3	625	13	US-10-027-632-44094	Sequence 44094, A
	7	43.8	3.3	625	13	US-10-027-632-44095	Sequence 44095, A
	8	43.8	3.3	625	14	US-10-027-632-44092	Sequence 44092, A
	9	43.8	3.3	625	14	US-10-027-632-44093	Sequence 44093, A
	10	43.8	3.3	625	14	US-10-027-632-44094	Sequence 44094, A
	11	43.8	3.3	625	14	US-10-027-632-44095	Sequence 44095, A
c	12	43.6	3.3	26320	13	US-10-034-650-13	Sequence 13, Appl
c	13	43.6	3.3	176373	14	US-10-095-407-17	Sequence 17, Appl
	14	43.4	3.3	409	11	US-09-918-995-6789	Sequence 6789, Ap
	15	43.4	3.3	13326	13	US-10-311-455-1685	Sequence 1685, Ap
c	16	43.2	3.3	1840	12	US-10-292-798-2003	Sequence 2003, Ap
c	17	43.2	3.3	1840	13	US-10-017-161-2359	Sequence 2359, Ap
c	18	43.2	3.3	18434	13	US-10-311-455-1979	Sequence 1979, Ap
	19	43.2	3.3	3673778	13	US-10-312-841-1	Sequence 1, Appli
c	20	43	3.3	201	9	US-09-864-761-18750	Sequence 18750, A
c	21	43	3.3	201	9	US-09-864-761-21744	Sequence 21744, A
c	22	43	3.3	395	9	US-09-864-761-5016	Sequence 5016, Ap
c	23	43	3.3	406	9	US-09-864-761-2003	Sequence 2003, Ap
	24	43	3.3	3066	13	US-10-148-687-8	Sequence 8, Appli
c	25	42.8	3.2	428	10	US-09-960-352-573	Sequence 573, App
	26	42.6	3.2	491	11	US-09-918-995-11962	Sequence 11962, A
	27	42.6	3.2	753	13	US-09-814-353-2034	Sequence 2034, Ap
	28	42.6	3.2	753	13	US-09-814-353-8380	Sequence 8380, Ap
	29	42.6	3.2	8530	13	US-10-311-455-1727	Sequence 1727, Ap
	30	42.4	3.2	143306	10	US-09-729-920-3	Sequence 3, Appli
	31	42.2	3.2	327	10	US-09-960-352-4630	Sequence 4630, Ap
c	32	42.2	3.2	7657	13	US-10-311-455-1995	Sequence 1995, Ap
c	33	42.2	3.2	7657	15	US-10-239-676-185	Sequence 185, App
c	34	42.2	3.2	12237	13	US-10-311-455-2331	Sequence 2331, Ap
	35	42.2	3.2	16509	13	US-10-311-455-1293	Sequence 1293, Ap
c	36	42	3.2	446	10	US-09-960-352-3400	Sequence 3400, Ap
c	37	42	3.2	1954	9	US-09-925-301-82	Sequence 82, Appl
c	38	42	3.2	2812	15	US-10-198-846-9693	Sequence 9693, Ap
	39	42	3.2	152331	14	US-10-095-407-16	Sequence 16, Appl
	40	41.8	3.2	642	11	US-09-866-050A-532	Sequence 532, App
	41	41.8	3.2	642	15	US-10-152-661-532	Sequence 532, App
c	42	41.8	3.2	11577	13	US-10-311-455-414	Sequence 414, App
	43	41.6	3.2	172	9	US-09-864-761-25620	Sequence 25620, A
	44	41.6	3.2	511	9	US-09-864-761-8959	Sequence 8959, Ap
c	45	41.6	3.2	6071	13	US-10-311-455-340	Sequence 340, App

ALIGNMENTS

RESULT 1
US-10-017-161-725
; Sequence 725, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 725
; LENGTH: 22684
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(22684)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(615)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (753)..(878)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17062)..(17263)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17606)..(17723)
; FEATURE:
; NAME/KEY: CDS

```
;   LOCATION: (19946)..(20099)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20577)..(20637)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22148)..(22484)
US-10-017-161-725

  Query Match             3.8%;  Score 49.6;  DB 13;  Length 22684;
  Best Local Similarity   61.7%;  Pred. No. 0.07;
  Matches   79;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        812 CCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATC 871
              ||  |  ||| ||||| || ||       || | ||  |   ||||| ||||| | ||||
Db      10903 CCTCCTCCTCCTCTTCATCCTCCTTCTTCTTCTATCATCATCATCATTATCATCATCATC 10962

Qy        872 ATCGCCATCCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTG 931
              |||  ||||  || |||||||||||||  |  |  ||||  ||| || |       | || ||
Db      10963 ATCATCATCATCATCATCATCATCATCACAGGGTCTCACTCTGTCACCCAGGCTGGAGTG 11022

Qy        932 TAGAGGAA 939
              || || |
Db      11023 CAGTGGCA 11030


RESULT 2
US-10-292-798-635
; Sequence 635, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 635
;   LENGTH: 29221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(29221)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(303)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (441)..(566)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15168)..(15253)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (16842)..(16951)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (17294)..(17411)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19634)..(19787)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20265)..(20325)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (21836)..(21908)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (26469)..(26554)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (26766)..(26901)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (27401)..(27442)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (28881)..(29021)
US-10-292-798-635

  Query Match             3.8%;  Score 49.6;  DB 12;  Length 29221;
  Best Local Similarity   61.7%;  Pred. No. 0.081;
  Matches   79;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        812 CCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATC 871
              ||  |  ||| ||||| || ||       || | ||  |   ||||| ||||| | ||||
Db      10591 CCTCCTCCTCCTCTTCATCCTCCTTCTTCTTCTATCATCATCATCATTATCATCATCATC 10650

Qy        872 ATCGCCATCCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTG 931
              |||  ||||  || |||||||||||||  |  |  ||||  ||| || |       | || ||
Db      10651 ATCATCATCATCATCATCATCATCATCACAGGGTCTCACTCTGTCACCCAGGCTGGAGTG 10710

Qy        932 TAGAGGAA 939
              || || |
Db      10711 CAGTGGCA 10718


RESULT 3
US-09-878-574-4292/c
; Sequence 4292, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 4292
;   LENGTH: 519
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(519)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3028-013-Q1-B1-D1
US-09-878-574-4292

  Query Match             3.4%;  Score 45;  DB 10;  Length 519;
  Best Local Similarity   47.7%;  Pred. No. 0.13;
  Matches  132;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
              | |||||| |  ||| ||| || ||||||||||    ||||  ||   |  ||   |||||
Db        511 AAAATTTAAAGTAAAAATAAAATATAAAAAATATGTAAATAAAAAAAATAAGATAAGAAAA 452

Qy        325 CGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCA 384
              | |      ||  | | |     |   | |  || ||     |   |||  | | | ||
Db        451 AAGAGAAAACAAATAAAAAAGTATAATCAATTAAAAAAAAAACAAGTAAAGATAAAATAA 392
```

```
Qy        385 CAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGG 444
Db        391 AAAAGTTAAAAAATATAATATAAAAAAAAATAAATTTGAATAAAAAAAATGATAAAATGA 332

Qy        445 AGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAAT 504
Db        331 ATTTAAGAAAAATATATATATAAAATAAAAAATAATAATAATAAATATATATTATTAAGAA 272

Qy        505 GCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTAT 541
Db        271 TAAAAATTATAATGGATAAATATTGAAAAGAAAATAT 235


RESULT 4
US-10-027-632-44092
; Sequence 44092, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44092
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44092

  Query Match             3.3%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
Db        469 GCAGTCAAATAGCTT 483


RESULT 5
US-10-027-632-44093
; Sequence 44093, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44093
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44093

  Query Match             3.3%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
Db        469 GCAGTCAAATAGCTT 483


RESULT 6
US-10-027-632-44094
; Sequence 44094, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
```

```
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44094
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44094

  Query Match             3.3%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | |  |  | | |  ||  |  ||||||| | |  |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    |  | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 7
US-10-027-632-44095
; Sequence 44095, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44095
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44095

  Query Match             3.3%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | |  |  | | |  ||  |  ||||||| | |  |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    |  | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 8
US-10-027-632-44092
; Sequence 44092, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44092
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44092

  Query Match             3.3%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | |  |  | | |  ||  |  ||||||| | |  |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    |  | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 9
US-10-027-632-44093
; Sequence 44093, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
```

```
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44093
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44093

  Query Match             3.3%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  || ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |   | | | ||  |  |||||||  | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 10
US-10-027-632-44094
; Sequence 44094, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44094
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44094

  Query Match             3.3%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  || ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |   | | | ||  |  |||||||  | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 11
US-10-027-632-44095
; Sequence 44095, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44095
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44095

  Query Match             3.3%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.31;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  || ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |   | | | ||  |  |||||||  | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
```

```
Db        469 GCAGTCAAATAGCTT 483
```

RESULT 12
US-10-034-650-13/c
; Sequence 13, Application US/10034650
; Publication No. US20030216558A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000128
; CURRENT APPLICATION NUMBER: US/10/034,650
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 09/474,377
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 26320
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-034-650-13

Query Match 3.3%; Score 43.6; DB 13; Length 26320;
Best Local Similarity 49.6%; Pred. No. 3.2;
Matches 112; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```
Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
Db       6178 AGAAGTAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAGAGAGAGGGGGGGAGAGA 6119

Qy        337 AATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTC 396
Db       6118 AAGAGAGAGAAAGAAAGAGAGAGAGAGGGGGAGAGAAAGAGAGAGAGAAAGAGAAAGA 6059

Qy        397 TGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAA 456
Db       6058 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGGAAGAAAGA 5999

Qy        457 AATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATA 502
Db       5998 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA 5953
```

RESULT 13
US-10-095-407-17/c
; Sequence 17, Application US/10095407
; Publication No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-17

Query Match 3.3%; Score 43.6; DB 14; Length 176373;
Best Local Similarity 48.8%; Pred. No. 9.6;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```
Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
Db     122513 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 122454

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
Db     122453 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 122394

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
Db     122393 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAAGAAAGAAAGAAAGCAAGCAAGCA 122334

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
Db     122333 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 122274

Qy        514 CA 515
Db     122273 AA 122272
```

RESULT 14
US-09-918-995-6789
; Sequence 6789, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6789
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(409)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-6789

Query Match 3.3%; Score 43.4; DB 11; Length 409;
Best Local Similarity 52.9%; Pred. No. 0.32;
Matches 92; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy        151 GCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGT 210
Db        164 GTTTGGGTTTGGGTTAGAAGGAGAATGAAATAAGAGAGATTGTATTTAAACTAGCCTAGG 223

Qy        211 ACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATT 270
Db        224 AAAATGGTGCACTGAAATAAAGGTTACTGCAGGGGTTTTCCCCATGACAAGCANAAAAAA 283

Qy        271 TATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
Db        284 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAGAAAAA 337
```

RESULT 15
US-10-311-455-1685
; Sequence 1685, Application US/10311455

; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1685
; LENGTH: 13326
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1685

Query Match 3.3%; Score 43.4; DB 13; Length 13326;
Best Local Similarity 53.2%; Pred. No. 2.4;
Matches 92; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy        282 TATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATAT 341
              | | |||  ||  | ||| |||||  |   ||| | |  | |  | | |||     ||| |
Db       9237 TTTTAAAATAAGGAGAAATAAATTATATGTTTGTATGGTAGAGGGATAGGGAGAGAATTT 9296

Qy        342 ACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGAC 401
              | |||| || | ||||   |||    |  | |  | | | |  | |  |||    |
Db       9297 TTAGAGAATAGAAATTTAGGGATGAAGTAGGAGGTGATAGTTATTAATAAATTAAGTGTT 9356

Qy        402 CTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTG 454
               | ||| |  ||| ||||  || | || ||   | |||  |   |||  ||||
Db       9357 ATAGAAGAAAAGAAAATGTAGGGATGAGATTGGGAAAGTATATTGTGGTTTTG 9409
```

Search completed: January 22, 2004, 01:20:32
Job time : 450.739 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 15:50:10 ; Search time 361.266 Seconds
(without alignments)
10468.497 Million cell updates/sec

Title: US-09-936-869-2
Perfect score: 1401
Sequence: 1 aaactcgacagcaaatatga.....ggaaatttgtgtacaaatc 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1401	100.0	1401	21	AAA99363	Plant PrAG1 promot
2		1320	94.2	1320	21	AAA99362	Plant PrAG1 promot
3	c	46.6	3.3	540	24	ABQ18618	Oligonucleotide fo
4		46.6	3.3	540	24	ABQ18619	Oligonucleotide fo
5	c	46.4	3.3	4492	23	ABL15002	Drosophila melanog
6	c	45.6	3.3	563	23	ABV50724	Human prostate exp
7	c	45	3.2	519	25	ABX22233	Human GDP-mannose
8		44.4	3.2	15923	24	ABL70179	Chemically treated
9		44.4	3.2	15923	24	AAS61132	Human gene regulat
10		44.4	3.2	15923	24	ABK31220	Signal transductio
11		44	3.1	6261	22	AAS46643	Tumour suppressor
12		44	3.1	6261	24	AAS61046	Human gene regulat
13		44	3.1	6261	24	ABK31160	Signal transductio
14	c	43.6	3.1	11084	20	AAX22302	Human IL-1ra BAC c
15		43.4	3.1	409	22	AAI92030	Human polynucleoti
16		43.4	3.1	13326	24	ABL33712	Human immune syste
17	c	43.2	3.1	18434	24	ABL34006	Human immune syste
18	c	43	3.1	201	22	ABA48463	Human breast cell
19	c	43	3.1	201	22	ABA51474	Human breast cell
20	c	43	3.1	201	22	ABA66368	Human foetal liver
21	c	43	3.1	201	22	ABA69502	Human foetal liver
22	c	43	3.1	201	22	ABA33430	Probe #11896 for g
23	c	43	3.1	201	22	ABA36424	Probe #14890 for g
24	c	43	3.1	201	22	AAK14785	Human brain expres
25	c	43	3.1	201	22	AAK17754	Human brain expres
26	c	43	3.1	201	22	AAK40526	Human bone marrow
27	c	43	3.1	201	22	AAK43573	Human bone marrow
28	c	43	3.1	201	22	AAI21285	Probe #11218 for g
29	c	43	3.1	201	22	AAI24360	Probe #14293 for g
30	c	43	3.1	201	22	AAI46560	Probe #15246 used
31	c	43	3.1	201	22	AAI49630	Probe #18316 used
32	c	43	3.1	201	22	AAI06989	Probe #6980 used t
33	c	43	3.1	201	22	AAI09899	Probe #9890 used t
34	c	43	3.1	201	23	ABS40089	Human liver single
35	c	43	3.1	201	23	ABS43218	Human liver single
36	c	43	3.1	201	24	ABS17726	Human genome-deriv
37	c	43	3.1	395	22	ABA46372	Human breast cell
38	c	43	3.1	395	22	ABA56938	Human foetal liver
39	c	43	3.1	395	22	ABA26550	Probe #5016 for ge
40	c	43	3.1	395	22	AAK05033	Human brain expres
41	c	43	3.1	395	22	AAK30563	Human bone marrow
42	c	43	3.1	395	22	AAI15180	Probe #5113 for ge
43	c	43	3.1	395	22	AAI36508	Probe #5194 used t
44	c	43	3.1	395	22	AAI04920	Probe #4911 used t
45	c	43	3.1	395	23	ABS30228	Human liver single

ALIGNMENTS

RESULT 1
AAA99363
ID AAA99363 standard; DNA; 1401 BP.
XX
AC AAA99363;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter DNA sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Pinus radiata.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART-) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;
XX
DR WPI; 2000-594442/56.
DR P-PSDB; AAB26795.
XX

PT Novel plant reproductive tissue promoter, useful to produce plants
PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 3; Fig 2; 51pp; English.
XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 1401 BP; 406 A; 281 C; 302 G; 412 T; 0 other;

Query Match 100.0%; Score 1401; DB 21; Length 1401;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60

Qy         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120

Qy        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180

Qy        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240

Qy        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300

Qy        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360

Qy        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420

Qy        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480

Qy        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540

Qy        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600

Qy        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660

Qy        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720

Qy        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780

Qy        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840

Qy        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900

Qy        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960

Qy        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020

Qy       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080

Qy       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCCACCTGAAGCCGT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCCACCTGAAGCCGT 1140

Qy       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200

Qy       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260

Qy       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320

Qy       1321 CTTTTGGACTGGTATTAGTAGTTGCAGCTTTGTTTTGCATACGCTGTGAGGATCTGTGCG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTTTTGGACTGGTATTAGTAGTTGCAGCTTTGTTTTGCATACGCTGTGAGGATCTGTGCG 1380

Qy       1381 CGGAAATTTTGTGTACAAATC 1401
              |||||||||||||||||||||
Db       1381 CGGAAATTTTGTGTACAAATC 1401
```

RESULT 2
AAA99362
ID AAA99362 standard; DNA; 1320 BP.
XX
AC AAA99362;
XX
DT 22-JAN-2001 (first entry)
XX
DE Plant PrAG1 promoter DNA sequence.
XX
KW Plant promoter; PrAG1; reproductive tissue; transgenic plant; cereal; ds.
XX
OS Pinus radiata.
XX
PN WO200055172-A1.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-NZ00031.
XX
PR 17-MAR-1999; 99NZ-0334715.
XX
PA (CART-) CARTER HOLT HARVEY LTD.
PA (TASM-) TASMAN BIOTECHNOLOGY LTD.
PA (UNMT) UNIV MICHIGAN TECHNOLOGICAL.
XX
PI Podila GK, Liu J, Karnosky DF;

XX
DR WPI; 2000-594442/56.
DR P-PSDB; AAB26795.
XX
PT Novel plant reproductive tissue promoter, useful to produce plants
PT which have a diminished reproductive capacity or which are sterile -
XX
PS Claim 2; Page 37-38; 51pp; English.
XX
CC This invention relates to a novel plant promoter gene. The promoter is
CC located in plant reproductive tissue, and the invention includes
CC transgenic plants containing the promoter. The promoter can be used to
CC produce plants which have a diminished reproductive capacity or which are
CC sterile. The constructs can also be used to transform agronomically
CC important plants in which modulation of reproductive capacity
CC (particularly the timing and abundance of flowering) is desirable,
CC e.g. cereals, rice, maize, wheat, barley, oats, rye, soybean and canola.
CC The present sequence represents DNA encoding the plant reproductive
CC promoter (PrAG1) of the invention.
XX
SQ Sequence 1320 BP; 390 A; 269 C; 280 G; 381 T; 0 other;

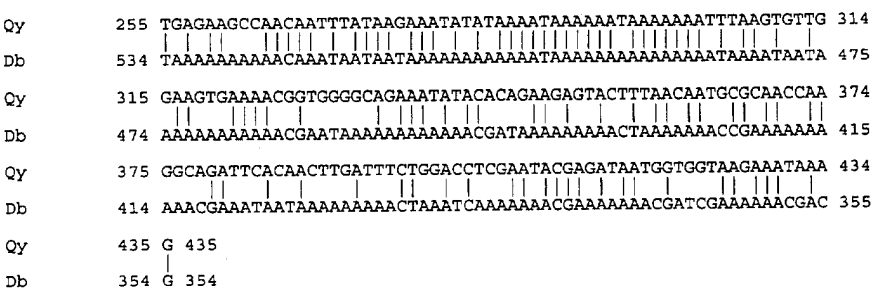
Query Match 94.2%; Score 1320; DB 21; Length 1320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAACTCGACAGCAAATATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCAT 60

Qy         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACATAACAAGCGGTGATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACT 120

Qy        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAA 180

Qy        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGAT 240

Qy        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAA 300

Qy        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAA 360

Qy        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGG 420

Qy        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATG 480

Qy        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGGACGAATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTA 540

Qy        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATGTGGGGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGA 600

Qy        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTA 660

Qy        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCCTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCC 720

Qy        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCCACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTC 780

Qy        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAA 840

Qy        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGAT 900

Qy        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTATCTCTATCTCTCCCTGGCAATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATAT 960

Qy        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATTGATCAACCTACCCGAAAAAACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAA 1020

Qy       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCTCTTATCCACTGTTCAATCATTCAGGTTTCTTCCCACTTTCAAGCAAAGGCGCCCGGA 1080

Qy       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTGGCCGTGTTCTTAGATTTTCAGGTACTTAAATGGACAATATTCCCCACCTGAAGCCGT 1140

Qy       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCTGAAAAAGATTTGTTTGTAGAAACAAACGATTGTAATATTTGCTTAAGTTGAGCTTAA 1200

Qy       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGGTTTGGTACCTAACTTGCCTTGTGGTTATTTGTTTCTCAGAACTCGGGCTGCGTCCA 1260

Qy       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTGTAGGAACGAACCAGCACAAGGGGTTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTG 1320
```

RESULT 3
ABQ18618/c
ID ABQ18618 standard; DNA; 540 BP.
XX
AC ABQ18618;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 5209.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.

XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 540 BP; 63 A; 40 C; 97 G; 340 T; 0 other;

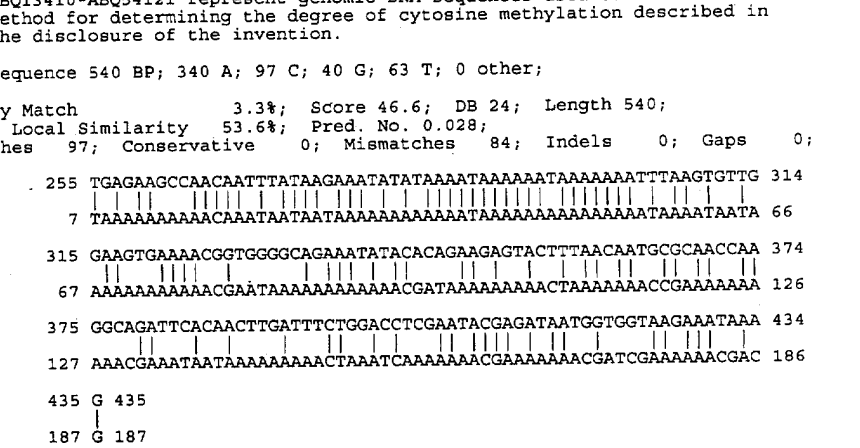
Query Match 3.3%; Score 46.6; DB 24; Length 540;
Best Local Similarity 53.6%; Pred. No. 0.028;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;



RESULT 4
ABQ18619
ID ABQ18619 standard; DNA; 540 BP.
XX
AC ABQ18619;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 5210.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.

XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 540 BP; 340 A; 97 C; 40 G; 63 T; 0 other;

Query Match 3.3%; Score 46.6; DB 24; Length 540;
Best Local Similarity 53.6%; Pred. No. 0.028;
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;



RESULT 5
ABL15002/c
ID ABL15002 standard; cDNA; 4492 BP.
XX
AC ABL15002;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39488.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX

OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB70899.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 39488; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4492 BP; 1158 A; 1245 C; 1141 G; 948 T; 0 other;

Query Match 3.3%; Score 46.4; DB 23; Length 4492;
Best Local Similarity 49.6%; Pred. No. 0.08;
Matches 119; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 804 TCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCA 863
Db 2878 TCCAGCGGCATGCTATCAATCTCATCGCCAACCACAACGATGCCATCGGGATCATCGTCA 2819

Qy 864 TTACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCA 923
Db 2818 TCTTCGTCCTCATCGCCGTCAGCCTCGTCGTCATCATCGTCGCCGGCCATGTGGATGCCA 2759

Qy 924 ATCGATTGTAGAGGAAAGGAAGAGGGAAGGGGCATATGTATTGATCAACCTACCCGAAAA 983
Db 2758 ACCTGCGGAAGAGCAACGGAAAATGTTATTGAATAATATGTTGCAAACTATAAACAAAGT 2699

Qy 984 AACAATCTGATCAGCCCTGCTCAATCTTGCTTATAAATCTCTTATCCACTGTTCAATCAT 1043
Db 2698 TAAAAAGCAAACAACCAAGCTTTAAAATATTAAGAATATGCATGTACTATGTTTTATCAT 2639

RESULT 6
ABV50724/c
ID ABV50724 standard; cDNA; 563 BP.
XX
AC ABV50724;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 50715.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 9866; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 563 BP; 45 A; 107 C; 22 G; 385 T; 4 other;

Query Match 3.3%; Score 45.6; DB 23; Length 563;
Best Local Similarity 49.2%; Pred. No. 0.053;
Matches 120; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 267 AATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACG 326
Db 520 AAGAAAAAAAAAAAAAAAAAAGGGAAAAGAAAAAAAAAAAAAAAGAGAAAAAAAAAAAAAG 461

Qy 327 GTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACA 386
Db 460 GACAGGAAAAAGGAAAAAAAAAAGAAGAAAAAAAACAAGAAGAAGAAAAAAAAAGAAAAAAA 401

Qy 387 ACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAG 446
Db 400 AAAAAAAAAAAAAAAAAAAAAAAGAAGGGAAAAAAAGAAAAAAAAAAAAAAAAAGAAAAAAA 341

Qy 447 TGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGC 506
Db 340 AAAAAAAAAAAAAAAAAGAAAGAAGAAAAAAAGAGAAGAAAAAAAAAAGAAAAAAAAAAA 281

Qy 507 AAGA 510
Db 280 AAAA 277

RESULT 7
ABX22233/c
ID ABX22233 standard; cDNA; 519 BP.
XX

AC ABX22233;
XX
DT 10-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #4290.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.
XX
OS Homo sapiens.
XX
PN US2002110548-A1.
XX
PD 15-AUG-2002.
XX
PF 11-JUN-2001; 2001US-0878574.
XX
PR 22-NOV-1996; 96US-0753233.
PR 03-DEC-1997; 97US-0984246.
PR 09-SEP-1998; 98US-0149674.
PR 14-JUN-1999; 99US-0333177.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Sullivan F, Kriz R, Kumar R;
XX
DR WPI; 2003-066673/06.
XX
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT rejection -
XX
PS Disclosure; SEQ ID NO 4292; 6pp; English.
XX
CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC human GM4,6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 519 BP; 129 A; 53 C; 23 G; 313 T; 1 other;

Query Match 3.2%; Score 45; DB 25; Length 519;
Best Local Similarity 47.7%; Pred. No. 0.074;
Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

Qy 265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
Db 511 AAAATTTAAAGTAAAAATAAAATATAAAAAATATGTAAATAAAAAAATAAGATAAGAAAA 452

Qy 325 CGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCA 384
Db 451 AAGAGAAAACAAATAAAAAAGTATAATCAATTAAAAAAAAAACAAGTAAAGATAAAATAA 392

Qy 385 CAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGG 444
Db 391 AAAAGTTAAAAAATATAATATAAAAAAAAAATAAATTTGAATAAAAAAAAATGATAAAATGA 332

Qy 445 AGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAAT 504
Db 331 ATTTAAGAAAAATATATATATAAAATAAAAAATAATAATAATAAATATATATATTAAGAA 272

Qy 505 GCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTAT 541
Db 271 TAAAAATTATAATGGATAAATATTGAAAAGAAAATAT 235

RESULT 8
ABL70179
ID ABL70179 standard; DNA; 15923 BP.
XX
AC ABL70179;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence#35.
XX
KW Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07471.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -
XX
PS Claim 1; SEQ ID NO 69; 24pp+sequence listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

Query Match 3.2%; Score 44.4; DB 24; Length 15923;
Best Local Similarity 50.8%; Pred. No. 0.47;
Matches 133; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

Qy 97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
Db 10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

```
Qy        157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
              || |         | ||| | | | |    |||||| | |||      | || |||||  ||
Db      10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy        217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
              ||    ||  | |      | ||     ||    || || |                |||| |
Db      10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945

Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
              |||| || | | |||||| |||| ||  || |||   |||||||   ||  |  | |
Db      10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy        337 AATATACACAGAAGAGTACTTT 358
              || |  ||     | | |||
Db      11006 AAAAATTACGTTGTATGATTTT 11027
```

RESULT 9
AAS61132
ID AAS61132 standard; DNA; 15923 BP.
XX
AC AAS61132;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #87.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 89; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

```
  Query Match             3.2%;  Score 44.4;  DB 24;  Length 15923;
  Best Local Similarity   50.8%;  Pred. No. 0.47;
  Matches  133;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy         97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
              | | | ||| |  | ||| ||| | |     |  | |||  |||    | || | |||
Db      10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

Qy        157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
              || |      | ||| | | | |    ||||| | |||      | || |||||  ||
Db      10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy        217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
              ||    ||  | |      | ||     ||    || || |                |||| |
Db      10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945

Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
              |||| || | | |||||| |||| ||  || |||   |||||||   ||  |  | |
Db      10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy        337 AATATACACAGAAGAGTACTTT 358
              || |  ||     | | |||
Db      11006 AAAAATTACGTTGTATGATTTT 11027
```

RESULT 10
ABK31220
ID ABK31220 standard; DNA; 15923 BP.
XX
AC ABK31220;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #32.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07472.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -

XX
PS Claim 1; SEQ ID No 63; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 15923 BP; 3153 A; 441 C; 4469 G; 7860 T; 0 other;

Query Match 3.2%; Score 44.4; DB 24; Length 15923;
Best Local Similarity 50.8%; Pred. No. 0.47;
Matches 133; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

```
Qy        97 GTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTGG 156
             | | | ||| |  | ||| ||| | |   |  | |||  |||    |  || |  |||
Db     10769 GAAATGGAGTATGGTTAGGGGATTTTAAGGATTTTTATTGTTTTTTGGAAATTTGTGTGG 10828

Qy       157 CTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTG 216
             || |        | ||| | | | |   |||||| | |||     | ||| |||||  ||
Db     10829 TTTATTTTAGTGACGAGTATTGATGTGTTAAAAGTTTTTTATAGTTTGTAAAGTATTTTT 10888

Qy       217 GATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAG 276
               ||   ||  | |      | ||     ||   || || |              ||||| |
Db     10889 ATTTTTGTTGTTGGTTAGATTGAGTTATTTATTTTGTTT---TTTTTTTTTTTTTATTAA 10945

Qy       277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
             |||| || | | | |||||| |||| ||  || |||   ||||||||    ||  |  | |
Db     10946 AAATTTAAATATTTAAAAATTAAAATATAAAAATGTGTTAAGTGAAAGAAGTTAGTTATA 11005

Qy       337 AATATACACAGAAGAGTACTTT 358
             || |  ||     |  | |||
Db     11006 AAAAATTACGTTGTATGATTTT 11027
```

RESULT 11
AAS46643
ID AAS46643 standard; DNA; 6261 BP.
XX
AC AAS46643;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #365.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
PS Claim 1; SEQ ID No 365; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

Query Match 3.1%; Score 44; DB 22; Length 6261;
Best Local Similarity 50.5%; Pred. No. 0.4;
Matches 107; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy       146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
             ||| |  |    || | | || |  |     | | | || | || |  | | ||  |
Db      1216 GGTCGTTTGTAGTTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy       206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
               | |  ||| | ||  |        |  ||        ||  ||| | | || |   | |
Db      1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGGATAGAGCGA 1335

Qy       266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
             | |||||  ||| ||| ||||||||||||||||||| | ||  ||  ||| ||  |
Db      1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy       326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
             |   | | |  |    || |||| |  ||
Db      1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427
```

RESULT 12
AAS61046
ID AAS61046 standard; DNA; 6261 BP.
XX
AC AAS61046;
XX
DT 29-JAN-2002 (first entry)

XX
DE Human gene regulation-associated gene oligonucleotide #1.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 1; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

Query Match 3.1%; Score 44; DB 24; Length 6261;
Best Local Similarity 50.5%; Pred. No. 0.4;
Matches 107; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
Db 1216 GGTCGTTTGTAGTTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy 206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
Db 1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGGATAGAGCGA 1335

Qy 266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
Db 1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy 326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
Db 1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427

RESULT 13
ABK31160
ID ABK31160 standard; DNA; 6261 BP.
XX
AC ABK31160;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #2.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07472.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
PS Claim 1; SEQ ID No 3; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 6261 BP; 1716 A; 190 C; 1696 G; 2659 T; 0 other;

Query Match 3.1%; Score 44; DB 24; Length 6261;
Best Local Similarity 50.5%; Pred. No. 0.4;
Matches 107; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy        146 GGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGT 205
              ||| | |   || | | || |  |  | | || || ||  |  | ||| |
Db       1216 GGTCGTTTGTAGTTTTAGTTATTCGGGAGGTTGAGGTAGGAGAATGGTGTGAATTCGGGA 1275

Qy        206 CAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAA 265
              | |  ||| | ||  |          ||  ||| | | || |   |  | |
Db       1276 GGCGGAGTTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTTTGGGGGATAGAGCGA 1335

Qy        266 CAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAAC 325
              | |||||  ||| ||| ||||||||||||||||||  | ||  ||  ||| || |
Db       1336 GATTTTATATTAAAAATAAAAAATAAAAAATAAAAAATTATATTTGAGGGATGTAGAGGA 1395

Qy        326 GGTGGGGCAGAAATATACACAGAAGAGTACTT 357
              |    | | |  |    ||  |||| |  ||
Db       1396 AGGTAAGTAAATTTTATAACGAAAGAATTATT 1427
```

RESULT 14
AAX22302/c
ID AAX22302 standard; DNA; 11084 BP.
XX
AC AAX22302;
XX
DT 22-JUN-1999 (first entry)
XX
DE Human IL-1ra BAC contiguous DNA sequence 95.
XX
KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX
OS Homo sapiens.
XX
PN WO9906426-A1.
XX
PD 11-FEB-1999.
XX
PF 03-AUG-1998; 98WO-US16102.
XX
PR 02-JUL-1998; 98US-0091650.
PR 04-AUG-1997; 97US-0054646.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI Pan Y;
XX
DR WPI; 1999-153692/13.
XX
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
XX
PS Example 5; Figure 4; 226pp; English.
XX
CC AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
CC sequences containing alternatively spliced forms of human IL-1ra. Such
CC fragments are used in the method of the invention which describes the
CC isolation of a novel human TANGO-77 encoding nucleic acid and protein.
CC Tango-77 is a member of the cytokine superfamily that is expected to
CC inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
CC may also bind to a new receptor so could regulate other cellular
CC processes associated with acute or chronic inflammation, e.g. asthma,
CC chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and
CC inflammatory bowel disease. It may also induce or suppress interleukins,
CC cytokines and growth factors. Modulators of this protein are used to
CC treat or prevent conditions associated with abnormal levels of
CC inflammation, or activity of IL-1 or its receptor complex.
XX
SQ Sequence 11084 BP; 3205 A; 2388 C; 2598 G; 2892 T; 1 other;

  Query Match   3.1%; Score 43.6; DB 20; Length 11084;
  Best Local Similarity 48.8%; Pred. No. 0.66;

  Matches 118; Conservative   0; Mismatches 124; Indels   0; Gaps   0;

```
Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
              |||| | | | |||| ||| ||  ||| ||  |||           | ||| | |
Db        780 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 721

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
              || || | | | |  | |  || || |  |  |  | | | | ||  ||
Db        720 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 661

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
                ||    |||  |||| || |  || | |||  |||| |||| | | |  |
Db        660 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 601

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
              | ||| ||| | ||  | |  || | |||| ||| |  |  | | || | ||||  |
Db        600 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 541

Qy        514 CA 515
              |
Db        540 AA 539
```

RESULT 15
AAI92030
ID AAI92030 standard; cDNA; 409 BP.
XX
AC AAI92030;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12090.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO12099.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 12090; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

```
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 409 BP; 183 A; 50 C; 77 G; 97 T; 2 other;

  Query Match             3.1%;  Score 43.4;  DB 22;  Length 409;
  Best Local Similarity   52.9%;  Pred. No. 0.18;
  Matches   92;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        151 GCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGT 210
              | |||| ||  || ||  | |||||  ||| |||  | |     || |  |||| | ||
Db        164 GTTTGGGTTTGGGTTAGAAGGAGAATGAAATAAGAGAGATTGTATTTAAACTAGCCTAGG 223

Qy        211 ACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATT 270
              | | |||    |      |   || |      |       | |   ||| |||| | ||
Db        224 AAAATGGTGCACTGAAATAAAGGTTACTGCAGGGGTTTTCCCCATGACAAGCANAAAAAA 283

Qy        271 TATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
               | || ||| | | |||| |||||| |||||||   ||     |  ||   ||||
Db        284 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAGAAAAA 337


Search completed: January 21, 2004, 17:31:47
Job time : 362.266 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:02:36 ; Search time 2794.45 Seconds
(without alignments)
12185.050 Million cell updates/sec

Title: US-09-936-869-2
Perfect score: 1401
Sequence: 1 aaactcgacagcaaatatga.....ggaaatttgtgtacaaatc 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	166	11.8	526	9	AW754659	AW754659 PC04G10 P
c	2	61.4	4.4	1101	29	CNS0039G	AL063921 Drosophil
	3	57.2	4.1	987	29	CNS00418	AL066537 Drosophil
	4	56.2	4.0	1201	13	BX461310	BX461310 BX461310
	5	55.2	3.9	1101	29	CNS003B2	AL064080 Drosophil
c	6	55.2	3.9	1101	29	CNS00DG1	AL069971 Drosophil
c	7	54.8	3.9	1041	9	AL559034	AL559034 AL559034
	8	54	3.9	964	29	CNS006N9	AL065781 Drosophil
	9	54	3.9	1101	29	CNS0174K	AL107582 Drosophil
c	10	54	3.9	1101	29	CNS017KE	AL108152 Drosophil
	11	53.4	3.8	1101	29	CNS0100X	AL098379 Drosophil
	12	52.8	3.8	997	29	CNS005TE	AL060767 Drosophil
	13	52.6	3.8	634	29	BX122696	BX122696 Danio rer
	14	52.6	3.8	1101	29	CNS0182P	AL108811 Drosophil
c	15	52.6	3.8	1114	13	BX407949	BX407949 BX407949
c	16	52.6	3.8	1201	9	AL566565	AL566565 AL566565
	17	52	3.7	783	29	CNS00AIS	AL055833 Drosophil
c	18	52	3.7	797	13	BX437575	BX437575 BX437575
	19	52	3.7	1101	29	CNS0181N	AL108773 Drosophil
c	20	51.8	3.7	623	28	AQ157491	AQ157491 nbxb0009H
c	21	51.8	3.7	830	29	CNS06P7Z	AL409029 T7 end of
c	22	51.8	3.7	1099	13	BX456575	BX456575 BX456575
c	23	51.6	3.7	946	13	BX416277	BX416277 BX416277
	24	51.6	3.7	1201	13	BX461128	BX461128 BX461128
c	25	51.4	3.7	1080	28	AQ738730	AQ738730 HS_5382_B
c	26	51.4	3.7	1101	29	CNS0039R	AL063932 Drosophil
c	27	51.2	3.7	997	29	CNS005TE	AL060767 Drosophil
c	28	51.2	3.7	1204	29	CNS016E2	AL106628 Drosophil
	29	51	3.6	673	29	CNS03XD5	AL264866 Tetraodon
c	30	50.8	3.6	1011	13	BX349844	BX349844 BX349844
c	31	50.6	3.6	929	13	BX408907	BX408907 BX408907
c	32	50.6	3.6	938	29	CNS006TJ	AL065906 Drosophil
c	33	50.4	3.6	626	28	AZ020346	AZ020346 RPCI-23-3
	34	50.4	3.6	924	29	CNS01GHN	AL143148 Anopheles
c	35	50.4	3.6	1001	29	CNS0064G	AL062781 Drosophil
c	36	50.2	3.6	1101	29	CNS017KX	AL108171 Drosophil
	37	50	3.6	1049	13	BX353130	BX353130 BX353130
	38	50	3.6	1101	29	CNS0106X	AL098595 Drosophil
c	39	50	3.6	1201	13	BX335607	BX335607 BX335607
	40	50	3.6	1201	13	BX385867	BX385867 BX385867
	41	49.6	3.5	658	28	AZ851742	AZ851742 2M0154E07
	42	49.6	3.5	735	29	AG095704	AG095704 Pan trogl
	43	49.6	3.5	1201	13	BX460090	BX460090 BX460090
	44	49.2	3.5	732	28	BZ146517	BZ146517 CH230-453
	45	49.2	3.5	760	29	AG056350	AG056350 Pan trogl

ALIGNMENTS

RESULT 1
AW754659
LOCUS AW754659 526 bp mRNA linear EST 01-MAY-2000
DEFINITION PC04G10 Pine TriplEx pollen cone library Pinus taeda cDNA clone PC04G10, mRNA sequence.
ACCESSION AW754659
VERSION AW754659.1 GI:7676379
KEYWORDS EST.
SOURCE Pinus taeda (loblolly pine)
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 526)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh ,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda TriplEx2 Sequencing Primer.
FEATURES Location/Qualifiers
source 1. .526

/organism="Pinus taeda"
/mol_type="mRNA"
/db_xref="taxon:3352"
/clone="PC04G10"
/dev_stage="immature"
/lab_host="E. coli BM25.8"
/clone_lib="Pine TriplEx pollen cone library"
/note="Organ: pollen cone; Vector: Lambda TriplEx; Site_1: SfiI (A); Site_2: SfiI (B); Immature pollen cones were collected in the early spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda TriplEx vector. Plasmid subclones in pTriplEx were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 134 a 99 c 128 g 146 t 19 others
ORIGIN

Query Match 11.8%; Score 166; DB 9; Length 526;
Best Local Similarity 97.1%; Pred. No. 3.1e-17;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1228 GTTATTTGTTTCTCAGAACTCGGGCTGCGTCCAACTGTAGGAACGAACCAGCACAAGGGG 1287
             ||| || | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 GTTCTTAGATTTTCAGAACTCGGGCTGCGTCCAACTGTAGGAACGAACCAGCACAAGGGG 213

Qy       1288 TTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTGCTTTTGGACTGGTATTAGTAGTTGCAG 1347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 TTGCAGCTTTTGCTGTTGCTGTTGCGCCCATTGCTTTTGGACTGGTATTAGTAGTTGCAG 273

Qy       1348 CTTTGTTTTGCATACGCTGTGAGGATCTGTGCGCGGAAATTTTGTGTACAAATC 1401
             ||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        274 CTTTGTTTTGCATACGCTGTGAGGATTTGTGCGCGGAAATTTTGTGTACAAATC 327
```

RESULT 2
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oscegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 4.4%; Score 61.4; DB 29; Length 1101;
Best Local Similarity 17.9%; Pred. No. 2;
Matches 140; Conservative 295; Mismatches 339; Indels 9; Gaps 1;

```
Qy         17 ATGATTTAGATTATGACCTAGAAATAAGCATAGCATTAAAGCATATACATAACAAGCGGT 76
              :| : | ::|::| ::    :::::|| |:   : :  ::  : ::  : : : : |::
Db       1013 WTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRKDGDDDKKGGKK 954

Qy         77 GATATACTCTGACTGCCACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTA 136
              :  |:| :     ::  : :::: ::||::: |::::| |:  :  : : |:::   |::
Db        953 KKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTKD 894

Qy        137 GTTTGGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTT 196
              ::::::::: :: :| : :  :     ::::: : : :: |::| |::::|::: :|: :
Db        893 DDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDAWAW 834

Qy        197 GACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTG 256
              :: : :|   :  :: :: ::| ::    ::  :::  ::::| :: ::   :|  | |
Db        833 KWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWTTTT 774

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              :: :  :  :|:|::||:: :::::::::: |  | ::::: : : :::::|:::: :
Db        773 TRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDAD 714

Qy        317 AGTGAAAACG~~~~~~~~~GTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCG 367
              |:: |::: :          |:|::  ::::: : | : : :| : :| ::|     :|:
Db        713 ADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDW 654

Qy        368 CAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAG 427
              :  :::| ::::     :| :: :|:: :: |  : : : :  ::: |:|  : ::|::
Db        653 DKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADR 594

Qy        428 AAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACG 487
              ::|:|:|  :|:::: : |: : ::||:     |  :|  | :    | :|:||  |
Db        593 DWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTTT 534

Qy        488 AATAAATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGG 547
              : | : |    | |  :   | : :  ::|     :| | :   | ||:  | | | |
Db        533 WTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAATTTTTTTTT 474

Qy        548 GGTCGGCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGATTGCAGT 607
              |    |   : |:      |    |: |     | :|      | |||  :||
Db        473 TTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTTWTTTWTAA 414

Qy        608 GTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATCCTCTC 667
                | ::   : || | ||  |    || :| :|||    ::     :| :|| :::
Db        413 AAAAAAWWWWWTWTATATKCCCCCCYCCCCCCYCCYCCCCTCCYYCCTCCYTCYCCCYYY 354

Qy        668 CATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTCCACTG 727
              : :: :: :| ||    ::: : |||  :   : : :::: |:|     |::::|| :|
Db        353 YYYYYYYYYTYCCCCTYYYYYYCCCTTTYYYYYYYCYYYYCCYTCNCNCCYYYYCCTYTY 294

Qy        728 CCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTTTCCCC 787
              |::   : |  |      | :|: :|  |:: :   :   : :| |||| || : :|::
Db        293 CYYYTTYTCCNCNCCYCCCCYCYYYCCYCYYCYCCCYYTYYCYTTTTTTYTTTYCYCYYT 234

Qy        788 TCC 790
              :|:
Db        233 YCY 231
```

RESULT 3
CNS00418
LOCUS CNS00418 987 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

```
            BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066537
VERSION     AL066537.1  GI:4942778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR09C16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      238 a    162 c     17 g    177 t    393 others
ORIGIN

  Query Match             4.1%;  Score 57.2;  DB 29;  Length 987;
  Best Local Similarity   27.1%;  Pred. No. 9.8;
  Matches  180;  Conservative   94;  Mismatches  386;  Indels    5;  Gaps    1;

Qy        258 GAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAA 317
              |||   :| ||    :|:  |      | ||  |||||  || |  | |       :|
Db         52 GAAAAAVAAAAAAANNDAWCGANANNNNANAAANAAAAANNNAATATATAANAAAAAAADA 111

Qy        318 GTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGC 377
                | | ||          | ||||| ||| | || |::          |   |   |  |
Db        112 ATAANAAAAAANAAAAAAAAATAAACAAAYAAAASHGNNNNNNGAGGGAANANNAANTNN 171

Qy        378 AGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGA 437
                                 |         ||||| | | ||    || |||  | |  |
Db        172 NNNAANNCTTNNANNGAANNAAANNNANAATACAANAAAAAAAAANAAAAAAAAAANAAAA 231

Qy        438 AGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAA 497
               | |   |   | | | ||  |              | | |  |    ||||| ||| ||
Db        232 AAAAAANAAAAAANTNAAGAANANAANATNNATATAAATAGNGAAAANAAAAAAAAAAAA 291

Qy        498 ATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCGGCATT 557
               | | ||  || |    |      |          ||| |  ||          :    :
Db        292 AAAAAAAAAAAAAAAANAAAANAANAANNNANAANNAAANNTNAATNNNCNCCSTAACKN 351

Qy        558 CACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTTGATTGCAGTGTGGGAGTTG 617
                |          |  | |   |        |                | |
Db        352 NANNNNNNAAANNNANCCANAGANNANNNAANNNNNNNAAAANNNNAAAAGANANNNAAA 411

Qy        618 CAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATCCTCTCCATTACTATT 677
                |  |    |    |    || : |:|    :|    :|  |::   :   :|   ::
Db        412 NANNAANNAANGANNANGNNATCYTCCYCCYTCYCCYTYYYCTTYYCTCYTCCYTCCCYY 471

Qy        678 ACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTCCACTGCCTGCACACG 737
               ::: : :|| |   ||   |     | : || : ||:: :    ::||:: |: ::: | |
Db        472 CYYYYYCYCTYTCTCTTCTTYTTYYTTCTYYCCYYYYTYTYYCTYYCCYYYYYYCYCYCC 531

Qy        738 TCTTAGTCAATCCATCTGCCTTCAAATAGGCATT-----ATTTTGTTCTTTCCCCTCCGA 792
               :|::     |    || :|   :|:  |    :    | |:      |||: ||:: |::|:: :
Db        532 YCYYCYCCCCCCYYCCYYCYCCYCCYCTCCCTYTCCYCYTTTYTTTYYCTYYCYYTYYY 591

Qy        793 CTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTG 852
               :|       || :    :|  :|    :: |:|    ::|:|:::: | :   : :: :|  ||
Db        592 YTCYYTCTCTYYTYYYCYYYCYCYYTCYCTCTYYCYTYYYYYTCTYCCYCYYCYCCTCTC 651

Qy        853 GATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATC 912
                :| :| || :: :|   : :: |:  :|:| :| :| :| :|              : :| ::
Db        652 YYYCCYCYTCYYYYYCTCYYYYTCYCCYCYCCYCCYCYYCCYCTCTCCCTCYCYCYTCYY 711

Qy        913 TCTCC 917
              |:|:|
Db        712 TYTYC 716


RESULT 4
BX461310
LOCUS       BX461310                1201 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX461310 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF025YA16 5-PRIME, mRNA sequence.
ACCESSION   BX461310
VERSION     BX461310.1  GI:31031291
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF025BA08QP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF025YA16"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      472 a    108 c    138 g    167 t    316 others
ORIGIN

  Query Match             4.0%;  Score 56.2;  DB 13;  Length 1201;
  Best Local Similarity   21.6%;  Pred. No. 13;
  Matches   81;  Conservative  139;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        141 GGTAAGGTTGGCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACT 200
                :: : ||  || :::: :: :::::  : |: :  ::||| ||  ||||:::::: :| :
Db        782 RRGRGGGGGGWWWDDDKKKKKKKKKKKTKKKKKDAAAAAAAAAAAAKDKDKDARADD 841

Qy        201 CTAGTCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAA 260
                  |:  ::: |    ::::  ::::: :::: ::: |    |:::::: ::::: :
```

```
Db        842 GGGGKKKKKKDDAAAAADDKKKKKKKKKKKKKKKKRAAAAWTTKKKKKKKKKKDKKKKK 901

Qy        261 GCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTG 320

Db        902 KKKKKKKDAAAAAAAAAAAAAADDKDDAAAAAAAAAAAAAAKKKKKKKKKKKKDAAAAA 961

Qy        321 AAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGA 380

Db        962 AAADKKKKKKKKKKDKKKDAAAAAAAAAAAAAAAAAAAAAADDKKKKKKKKKKKKKK 1021

Qy        381 TTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGA 440

Db       1022 KKKKKKKKKDAADKKKKKKKAAAAAAAAAAAAAAKKDAAADDKKKKKDAAAAADKKKKK 1081

Qy        441 GTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATA 500

Db       1082 AAAAAAAADKDDDAAAAAKKKKKKKKKDAAAAAADDDKDKKKKKKDDKKKAAAAAAAA 1141

Qy        501 TAATGCAAGAGTGCA 515

Db       1142 AAAAAAKAAAKKKAA 1156
```

RESULT 5
CNS003B2
LOCUS CNS003B2 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR08A08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL064080
VERSION AL064080.1 GI:4941836
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08A08"
/clone_lib="RPCI-98"
/note="end : T7"
BASE COUNT 399 a 197 c 187 g 196 t 122 others
ORIGIN

Query Match 3.9%; Score 55.2; DB 29; Length 1101;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 117; Conservative 17; Mismatches 123; Indels 0; Gaps 0;

```
Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316

Db        844 AAAAAWGAAAAAATTRTATKRAAAAAAAWAAAAAAAAARAAAGAAAGAGGRAAAAATGAA 903

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376

Db        904 AAAAAAAAAAATAAAAAAAAAAAAAAWTAAAAAAAAAAAAAAAAAAWAWNGAAAAAGAGAAA 963

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436

Db        964 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAAANGAAAATGAGT 1023

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496

Db       1024 GAKTGKGAAKAAWAAATGAARATGANGTAAGARAGAAAAAAAAAAAAGARAGAAATGGAARA 1083

Qy        497 AATATAATGCAAGAGTG 513

Db       1084 AAAAAAAAAAAAGWGKG 1100
```

RESULT 6
CNS00DG1/c
LOCUS CNS00DG1 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR27J03 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069971
VERSION AL069971.1 GI:4947993
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27J03"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 227 a 174 c 33 g 361 t 306 others
ORIGIN

Query Match 3.9%; Score 55.2; DB 29; Length 1101;
Best Local Similarity 34.3%; Pred. No. 20;
Matches 114; Conservative 62; Mismatches 156; Indels 0; Gaps 0;

```
Qy        173 GAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATTGGATTGCCTTTGTCGGG 232
```

```
Db        952 KAGGWAGAAAARGARAAAAAARARRAGGAKTGGAAAGGTGGAARGRGATTGKTAAAGRGR 893

Qy        233 GCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAA 292

Db        892 GGATRAAAWVTTTWTSWAAAAAVDTSSTSAAAAASAAAAVARAAAAAAAAAAAAAAAAAA 833

Qy        293 AAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAG 352

Db        832 WAAAAAAATWAAAAAASTSKTAAAAAASSSAVAASAWWWATGSSRATAAAWRARSARGRA 773

Qy        353 TACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGA 412

Db        772 AARTWTAAAAAAAAGAARAAAAAASSVASASWAWRAVSSASAAAWSAAWAAARVAAAATT 713

Qy        413 GATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGC 472

Db        712 GATAARGGTGGGAGAAAARTTWWAATTAAWGWWRRAATGRTAWRARARAATATGAARSAG 653

Qy        473 ACAAAATGGAGGACGAATAAATGAAATATAAT 504

Db        652 ARTWAATWRAAAATGWAAAATTTARTTTGAAT 621


RESULT 7
AL559034/c
LOCUS       AL559034                1041 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL559034 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ010YL15 5-PRIME, mRNA sequence.
ACCESSION   AL559034
VERSION     AL559034.2  GI:31283167
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12904134.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            8934.f For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DJ010CF08QP1&cluster=8934.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DJ010CF08QP1.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ010YL15"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      153 a    190 c    201 g    335 t    162 others
ORIGIN

  Query Match             3.9%;  Score 54.8;  DB 9;  Length 1041;
  Best Local Similarity   33.7%;  Pred. No. 23;
  Matches   86;  Conservative   29;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        254 GTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTT 313

Db        460 GAGTRWAAAAARGRCCTGWRAAAAAAAAAAARAATATAAGAAATGAAAAAAARWAAAAGGAA 401

Qy        314 GGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCA 373

Db        400 AAAARAAAAWAAAWGRARRARRAAAAWAAAAAAARAAAAWAAWAAAAARAWARAAAAAAA 341

Qy        374 AGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAA 433

Db        340 NNAAAAANNNNNNANNANNANNNNNNNNANNAAAAAAANNNNANNNNGGAANNNAANNA 281

Qy        434 AGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAA 493

Db        280 ANNAANNANNNNNNNANNNNNNNANNANNTGAGGGCGRRARMCCGRGCARAGRAGCCAR 221

Qy        494 TGAAATATAATGCAA 508

Db        220 RCARATCCAACGARM 206


RESULT 8
CNS006N9
LOCUS       CNS006N9                964 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065781
VERSION     AL065781.1  GI:4944661
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14J19"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      409 a    111 c     71 g    130 t    243 others
ORIGIN

  Query Match             3.9%;  Score 54;  DB 29;  Length 964;
  Best Local Similarity   34.0%;  Pred. No. 32;
  Matches   86;  Conservative   64;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        264 AACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAA 323
```

```
Db        664 AAAAATTGTAGTWAWTAKTAAAAAATATAAAAAADAAADAADAAAKKTKWDKAADAAAAA 723

Qy        324 ACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTC 383

Db        724 AAGWWAKKDAAAAAAAAAAAAAAAKAKGTATADADAAAAADTAKAKWDWADAKDDDKKAGA 783

Qy        384 ACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTG 443

Db        784 AKADKWARKAAARAAAAWCKKDARKAKAWAAKARWAKAGTGWAAKARAGKATAGAGAGAK 843

Qy        444 GAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db        844 AAKAKDKDKAGKDAKADAAKAKADAKAKADKADDAKAGADKAGAKAKKKGKANAGDAKAK 903

Qy        504 TGCAAGAGTGCAT 516

Db        904 AKAAAAAAARWAK 916
```

RESULT 9
CNS0174K
LOCUS CNS0174K 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN17P20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL107582
VERSION AL107582.1 GI:5627886
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN17P20"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 365 a 90 c 235 g 206 t 205 others
ORIGIN

Query Match 3.9%; Score 54; DB 29; Length 1101;
Best Local Similarity 42.5%; Pred. No. 31;
Matches 147; Conservative 22; Mismatches 177; Indels 0; Gaps 0;

```
Qy        253 TGTGAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGT 312

Db        216 TATAATAGRATAATAATAATTAGGGWTAAAATAAAAAAAAWATAAAGAAATAAAAAARGAA 275

Qy        313 TGGAAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACC 372

Db        276 AATTAAAAAAAAAAARGWAGAAAAATATTAAAAAAAAGTARWTATTWAAAAATAATGAAAT 335

Qy        373 AAGGCAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATA 432
```

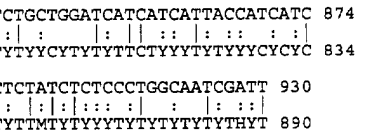
```
Db        336 AAWTTATATWTAAAAAARGAAAGATGRAWAGRAAAATRAAAAWAAAAATRATAAAAAATA 395

Qy        433 AAGGAAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAA 492

Db        396 AWTWAAAAAAAWWATTAATATAAATAAATAAAAAAAAAGTTAAATATAAAATTAAAAAAT 455

Qy        493 ATGAAATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCG 552

Db        456 AAGTATATAAAAAAAATWGAGTTTGATTSGAATAAAGTGATAGATGATTKAGAGGCGGAG 515

Qy        553 GCATTCACATGGGCGTCGCATTCAGGGGGTGTCATAGCGGTCCTTT 598

Db        516 GCAMTCAMGAASGTCGCKAAAWAGCGGTCACCGGTATCSATTCTYT 561
```

RESULT 10
CNS017KE/c
LOCUS CNS017KE 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN37F11 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108152
VERSION AL108152.1 GI:5628456
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F11"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
BASE COUNT 302 a 83 c 32 g 379 t 305 others
ORIGIN

Query Match 3.9%; Score 54; DB 29; Length 1101;
Best Local Similarity 35.1%; Pred. No. 31;
Matches 99; Conservative 57; Mismatches 126; Indels 0; Gaps 0;

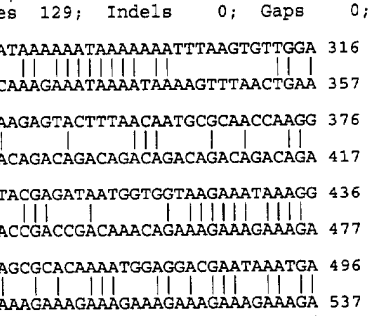
```
Qy        264 AACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAA 323

Db        711 MAGDAAKKGAAAAAAAAAAAAHDADAAAAAAAAAAAGGWAAAAKKAKKKKKMAGRAKAKACA 652

Qy        324 ACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTC 383

Db        651 AAKKAAGKTCAGACACKAKAAAAAAGDCKKMGGKDAKAAARAKMKAAAKKAKKAKKKAKK 592

Qy        384 ACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTG 443

Db        591 KKKAKKAAAKKDWCRAAKKWAAAAKDKKAGAGAADKKGKAAAAGAAKAKAAATATAWSGA 532

Qy        444 GAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503
```

```
Db        531 TTGKWCAGKATAAGAWKAAKAAAAAKARGKAAKWATCKAKAAKKAKAAAAKATGAATATA 472

Qy        504 TGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGT 545

Db        471 ATTTAATGTTTTTGTTATTATTTTTTTTTATTTATTATATAT 430


RESULT 11
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      195 a    108 c    131 g    161 t    506 others
ORIGIN

  Query Match             3.8%;  Score 53.4;  DB 29;  Length 1101;
  Best Local Similarity   18.2%;  Pred. No. 39;
  Matches   82;  Conservative  187;  Mismatches  179;  Indels    3;  Gaps    1;

Qy         96 TGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGCTTG 155

Db        632 KRGWGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWGWKDKWTDKKRDGKRGDGDRKDKGTG 691

Qy        156 GCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTACATT 215

Db        692 GDGDSKAKGDRAWAWRAKATRAAAADATAAATAAKATAWATKTTTKKTTTTTTTKTTTTT 751

Qy        216 GGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTATAA 275

Db        752 TTTTTTKKTTTARKADRDDDDKAKRDGDKDGDWDDADKDWKAGRDRDDDTRKWGWKDRAG 811

Qy        276 GAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAG 335

Db        812 KKKKAKKKDHKTAWWWADDTGTAWATWAKTATAKTGKAKATKDBKRGRGGGGRGDGRGAG 871

Qy        336 ---AAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392

Db        872 RRKRARGRKRGRRRKDKRKRKKRTKKAGDGDTGAGTDKAGDKKADGDGWGWKRKWDSKT 931

Qy        393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452

Db        932 WKTTMTGTDDTTTWKGRNRRGRGRAGAGRRRGRKGKAGKGAGGKWWKWAWADABWTWTWD 991

Qy        453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGT 512

Db        992 WKDWKWKWTDWGWGKGGWGDGRKGGAKDGAKRNRKNRDGMRYGNRAWKWMDDAATAKARW 1051

Qy        513 GCATTTCCCTATTATTTCCAGAAATGTATAT 543

Db       1052 ADDDWDADYDDDADADYWKWKATWMDBWCDT 1082


RESULT 12
CNS005TE
LOCUS       CNS005TE                997 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION     AL060767.1  GI:4943573
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12K22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       89 a     99 c     13 g    258 t    538 others
ORIGIN

  Query Match             3.8%;  Score 52.8;  DB 29;  Length 997;
  Best Local Similarity   20.6%;  Pred. No. 50;
  Matches   61;  Conservative  111;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        635 TCCATTCATCCCAAAACCTAAATTTATCCTCTCCATTACTATTACCTACACCTATACCTA 694

Db        595 YCYYTCYTYYCYYYCYCTYYYYCYTYCYYYYCYYCYYYYYTCYTYYYTTCTCYYTYYYYY 654

Qy        695 GTAAATATGTCCTGCCTTGTAACTCCTCCACTGCCTGCACACGTCTTAGTCAATCCATCT 754

Db        655 TYYAYKYYCYYYYYTCTCYYYCYCYCYYYTCTCYCYYCYYYYTTCTCTYTTYYTYYYTYY 714

Qy        755 GCCTTCAAATAGGCATTATTTTGTTCTTTCCCCTCCGACTGAAAGGCTATCGACCGACCG 814

Db        715 TYTCTCYYYYYYYYYYYTYYTTTCTYYTYYYYTTTYTCTYTYTYTYYTTYYYYCYYYCCYC 774
```



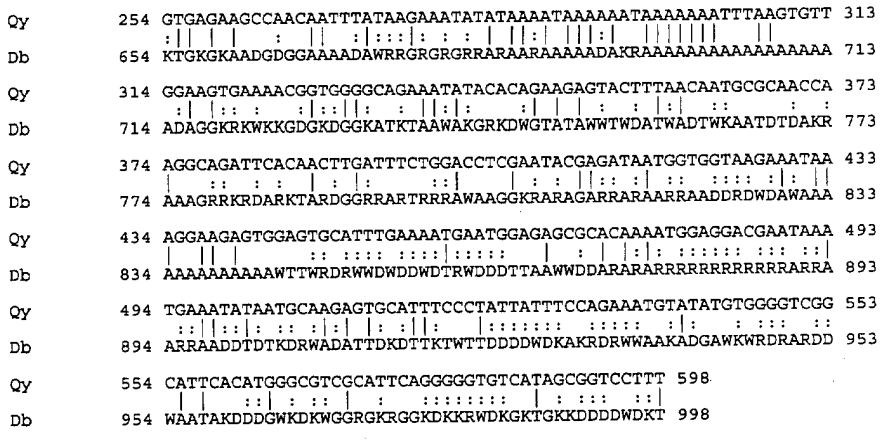
RESULT 13
BX122696
LOCUS BX122696 634 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-58A23, genomic survey sequence.
ACCESSION BX122696
VERSION BX122696.1 GI:27953626
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 634)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the SP6 end of BAC 58A23. 58A23 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES Location/Qualifiers
source 1. .634
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-58A23"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
BASE COUNT 363 a 75 c 151 g 45 t
ORIGIN

Query Match 3.8%; Score 52.6; DB 29; Length 634;
Best Local Similarity 50.2%; Pred. No. 63;
Matches 130; Conservative 0; Mismatches 129; Indels 0; Gaps 0;



RESULT 14
CNS0182P
LOCUS CNS0182P 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108811
VERSION AL108811.1 GI:5629115
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37D10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

Query Match 3.8%; Score 52.6; DB 29; Length 1101;
Best Local Similarity 21.7%; Pred. No. 52;
Matches 75; Conservative 132; Mismatches 138; Indels 0; Gaps 0;



RESULT 15
BX407949/c
LOCUS BX407949 1114 bp mRNA linear EST 15-MAY-2003
DEFINITION BX407949 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CS0DM001YI02 3-PRIME, mRNA sequence.
ACCESSION BX407949
VERSION BX407949.1 GI:30768711
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1114)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6978.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AS014ZE09NP1&cluster=6978.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AS014ZE09NP1.
FEATURES              Location/Qualifiers
     source           1. .1114
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DM001YI02"
                      /tissue_type="FETAL LIVER"
                      /dev_stage="fetal"
                      /clone_lib="Homo sapiens FETAL LIVER"
                      /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                      was primed with a NotI-oligo(dT) primer. Five prime end
                      enriched, double-strand cDNA was digested with Not I and
                      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                      vector. Library was not normalized."
BASE COUNT      139 a    161 c     51 g    670 t     93 others
ORIGIN

  Query Match             3.8%;  Score 52.6;  DB 13;  Length 1114;
  Best Local Similarity   44.2%;  Pred. No. 52;
  Matches  144;  Conservative   17;  Mismatches  164;  Indels    1;  Gaps    1;

Qy        257 AGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGA 316
              |:||   ||:||   : |||||: | | |||| |||||| ||||||||   ||   ||| |
Db        399 ARAAAAAAAMAAMCAMAAAGAARAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAATTGAA 340

Qy        317 AGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGG 376
              || |:|  | |||    |    | |    ::   |:         ||   | ||  :|
Db        339 AGGGRAGGCNGTGTAAGANTTGTTTNTGGRSGGNASAGTGNNTTAAAAAAGAAAAARAAA 280

Qy        377 CAGATTCACAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGG 436
              | |   | ||   ||    ||       ||       | |  ||||||||   || ||| |
Db        279 AAAAAAAAAAAAAAGAAAAGTGTNTTGAGATAGNNNGGTNNTGGTGGTATTTAAAAAAAG 220

Qy        437 AAGAGTGGAGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGA 496
              :| |   :        |:||| :|| || | |   | |||| | | |   || ||| ||
Db        219 MAAAAAAAMAAAAGAAAGRAAAARAAAGG-GGGAAGAAAAAAGGAAAGGAAAAAAAAAGA 161

Qy        497 AATATAATGCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTATATGTGGGGTCGGCAT 556
              |: | |||   | || | |       |  |      | | |||   | |   | |
Db        160 ARGAAAATAAGAAAGGGAAAAAAAAAAAAAAARAAAAANATGGGGAAGGTAGTTTTTTCA 101

Qy        557 TCACATGGGCGTCGCATTCAGGGGGT 582
              ||  | : |  ||  || |||||::|
Db        100 TCTAAARTGTCTCTGATGCAGGGKRT 75


Search completed: January 22, 2004, 00:54:16
Job time : 2795.45 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 16:52:57 ; Search time 94.7682 Seconds
(without alignments)
6525.161 Million cell updates/sec

Title: US-09-936-869-2
Perfect score: 1401
Sequence: 1 aaactcgacagcaaatatga.....ggaaattttgtgtacaaatc 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	73.4	5.2	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	61	4.4	7218	1	US-08-232-463-14	Sequence 14, Appl
c	3	43.6	3.1	176373	3	US-09-128-155-17	Sequence 17, Appl
	4	42	3.0	152331	3	US-09-128-155-16	Sequence 16, Appl
c	5	41	2.9	2255	2	US-08-741-134-1	Sequence 1, Appli
c	6	40	2.9	87350	3	US-08-781-891-79	Sequence 79, Appl
c	7	40	2.9	87350	4	US-09-618-166-79	Sequence 79, Appl
c	8	40	2.9	87543	4	US-09-791-211-3	Sequence 3, Appli
c	9	39.4	2.8	435	4	US-09-397-787-171	Sequence 171, App
c	10	39.4	2.8	1024	4	US-09-328-475C-64	Sequence 64, Appl
	11	39.2	2.8	1753	3	US-09-058-260-21	Sequence 21, Appl
	12	39.2	2.8	1756	3	US-09-058-260-31	Sequence 31, Appl
	13	39.2	2.8	1776	3	US-09-058-260-23	Sequence 23, Appl
	14	39.2	2.8	1952	3	US-09-058-260-5	Sequence 5, Appli
	15	39.2	2.8	4090	2	US-08-781-802-5	Sequence 5, Appli
	16	39.2	2.8	4090	3	US-08-694-078-5	Sequence 5, Appli
c	17	38.8	2.8	3564	2	US-07-594-921C-9	Sequence 9, Appli
c	18	38.8	2.8	3564	4	US-09-197-948-9	Sequence 9, Appli
c	19	38.6	2.8	202001	4	US-09-734-674-3	Sequence 3, Appli
	20	38.2	2.7	289	3	US-09-007-005-17	Sequence 17, Appl
	21	38.2	2.7	289	3	US-09-244-796-17	Sequence 17, Appl
	22	37.6	2.7	1896	3	US-09-058-260-3	Sequence 3, Appli
	23	37.6	2.7	6263	2	US-08-781-802-3	Sequence 3, Appli
	24	37.6	2.7	6263	3	US-08-694-078-3	Sequence 3, Appli
	25	37.4	2.7	475	4	US-09-702-705-36	Sequence 36, Appl
	26	37.4	2.7	475	4	US-09-736-457-36	Sequence 36, Appl
c	27	37.4	2.7	4235	4	US-09-702-705-317	Sequence 317, App
c	28	37.4	2.7	4235	4	US-09-736-457-317	Sequence 317, App
c	29	37.4	2.7	72604	4	US-09-268-992-7	Sequence 7, Appli
c	30	37.4	2.7	72604	4	US-09-657-474-7	Sequence 7, Appli
c	31	37	2.6	5720	4	US-09-442-100-1	Sequence 1, Appli
c	32	37	2.6	5720	4	US-08-939-106-1	Sequence 1, Appli
c	33	36.8	2.6	289	3	US-09-007-005-17	Sequence 17, Appl
c	34	36.8	2.6	289	3	US-09-244-796-17	Sequence 17, Appl
	35	36.8	2.6	246240	2	US-08-724-394A-20	Sequence 20, Appl
	36	36.8	2.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
	37	36.8	2.6	246240	2	US-08-724-394A-22	Sequence 22, Appl
	38	36.6	2.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	39	36.4	2.6	658	3	US-08-998-416-595	Sequence 595, App
c	40	36.4	2.6	1164	4	US-09-538-871-3	Sequence 3, Appli
c	41	36.4	2.6	1232	4	US-09-538-871-1	Sequence 1, Appli
c	42	36.2	2.6	2013	4	US-09-134-001C-103	Sequence 103, App
	43	36.2	2.6	176373	3	US-09-128-155-17	Sequence 17, Appl
c	44	36	2.6	2518	3	US-09-433-699-3	Sequence 3, Appli
c	45	36	2.6	1230025	4	US-09-198-452A-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match 5.2%;  Score 73.4;  DB 1;  Length 7218;
```

```
  Best Local Similarity   6.2%;  Pred. No. 8.5e-11;
  Matches   26;  Conservative  236;  Mismatches  157;  Indels    0;  Gaps    0;

Qy          93 CACTGTACTTGAGGAAAGGTAGTGGACTCTGCTCAGGTACATTAGTTTGGTAAGGTTGGC 152
               ||||||| ||     |    |    |    |    || || |  | ||||||| ::  ::
Db        1487 CACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRR 1428

Qy         153 TTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGTAC 212
                 ::     ::: :: : :::::: ::::::: ::::::  :   ::    ::  ::: :
Db        1427 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1368

Qy         213 ATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATTTA 272
               :  :::  :     :  ::::   ::: ::   :::   : : ::::::  :: ::   :
Db        1367 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1308

Qy         273 TAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGG 332
                ::::::  : : :::: :::::: :::::::   ::: :  ::::: ::::: :: ::::
Db        1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248

Qy         333 CAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGA 392
                ::::: : : : ::::::: :    :: :: : : ::  :::: :::   : ::   ::
Db        1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188

Qy         393 TTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATT 452
                  :::    ::: : :::: :: :: :: :::::: :::::::::: :::: : :
Db        1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128

Qy         453 TGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAG 511
                :::::  ::: :::::: : : :::: :::::: ::: ::: :::: : :: : :::::
Db        1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069


RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.4%;  Score 61;  DB 1;  Length 7218;
  Best Local Similarity   2.2%;  Pred. No. 3e-07;
  Matches    7;  Conservative  200;  Mismatches  110;  Indels    0;  Gaps    0;

Qy         603 GCAGTGTGGGAGTTGCAACATGTACCAACAAATCCATTCATCCCAAAACCTAAATTTATC 662
               | ||  || || :: :  : : : :: :  ::: ::: :::: :::  ::: ::
Db        1055 GGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114

Qy         663 CTCTCCATTACTATTACCTACACCTATACCTAGTAAATATGTCCTGCCTTGTAACTCCTC 722
               :::::: :: :: :: ::: : ::: : ::: :  :   : : :::: :::: :  ::::::
Db        1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

Qy         723 CACTGCCTGCACACGTCTTAGTCAATCCATCTGCCTTCAAATAGGCATTATTTTGTTCTT 782
               : :: ::: : : : ::::  ::  ::: ::: :::::   :   : :: :::: :::::
Db        1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

Qy         783 TCCCCTCCGACTGAAAGGCTATCGACCGACCGACCGCTCATCTTCTTCTTCTGCGCAATT 842
               :::::::  ::      :: ::  ::  ::  :: ::: ::::::::::: : :  ::
Db        1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294

Qy         843 TTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCACCATCATCATCATGATGG 902
               :::::: ::   :: :: :: :: :: :: :: :: ::::: :: :: :: :: :  :
Db        1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354

Qy         903 TATCTCTATCTCTCCCT 919
               : ::::: :::::::::
Db        1355 YYYYYYYYYYYYYYYYY 1371


RESULT 3
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

  Query Match             3.1%;  Score 43.6;  DB 3;  Length 176373;
  Best Local Similarity  48.8%;  Pred. No. 0.15;
  Matches  118;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
               |||| | | | |||| ||| ||  ||| ||   |||          | |||  | |
Db     122513 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 122454

Qy         334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
```

```
Db     122453 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 122394

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453

Db     122393 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 122334

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513

Db     122333 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 122274

Qy        514 CA 515

Db     122273 AA 122272


RESULT 4
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             3.0%;  Score 42;  DB 3;  Length 152331;
  Best Local Similarity   48.3%;  Pred. No. 0.4;
  Matches  117;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333

Db     146880 AGGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 146939

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393

Db     146940 GAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 146999

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453

Db     147000 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 147059

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513

Db     147060 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGAAAAAGAAAGAAAGAAAGAAAG 147119

Qy        514 CA 515

Db     147120 AA 147121


RESULT 5
US-08-741-134-1/c
; Sequence 1, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,163
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2255 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  160..1398
US-08-741-134-1

  Query Match             2.9%;  Score 41;  DB 2;  Length 2255;
  Best Local Similarity   63.9%;  Pred. No. 0.089;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy        820 TCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATCATCGCCAT 879

Db        826 TCTTCTTCTTCTTGGGTGCTTCCTCCTCTTCTTCATCGTCCTCATCATCATCTTCTTCAT 767

Qy        880 CCCCACCATCATCATCATGATGGTATCTCTATCTCTC 916

Db        766 CCTCTTCGTCTTCATCATCATCCTCTTCATCACTGTC 730


RESULT 6
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
```

```
;     ADDRESSEE:  SEED and BERRY LLP
;     STREET:  6300 Columbia Center, 701 Fifth Avenue
;     CITY:  Seattle
;     STATE:  Washington
;     COUNTRY:  USA
;     ZIP:  98104-7092
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/781,891
;     FILING DATE:  27-DEC-1996
;     CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  No. 6090620tenburg Ph.D., Carol
;     REGISTRATION NUMBER:  39,317
;     REFERENCE/DOCKET NUMBER:  240052.419
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (206) 622-4900
;     TELEFAX:  (206) 682-6031
; INFORMATION FOR SEQ ID NO:  79:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  87350 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-08-781-891-79

  Query Match             2.9%;  Score 40;  DB 3;  Length 87350;
  Best Local Similarity   50.8%;  Pred. No. 1.1;
  Matches  120;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328

Db      49802 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49743

Qy        329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387

Db      49742 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49683

Qy        388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447

Db      49682 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49623

Qy        448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db      49622 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49567

RESULT 7
US-09-618-166-79/c
; Sequence 79, Application US/09618166
; Patent No. 6583112
;  GENERAL INFORMATION:
;       APPLICANT: Fu, Ying-Hui
;                  Yu, Chang-En
;                  Oshima, Junko
;                  Mulligan, John T.
;                  Schellenberg, Gerald D.
;       TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                           WERNER'S SYNDROME
;       NUMBER OF SEQUENCES: 209
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/618,166
;            FILING DATE: 17-Jul-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 240052.419C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 87350 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-618-166-79

  Query Match             2.9%;  Score 40;  DB 4;  Length 87350;
  Best Local Similarity   50.8%;  Pred. No. 1.1;
  Matches  120;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328

Db      49802 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49743

Qy        329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387

Db      49742 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49683

Qy        388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447

Db      49682 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49623

Qy        448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503

Db      49622 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49567

RESULT 8
US-09-791-211-3/c
; Sequence 3, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 3
;   LENGTH: 87543
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 7421
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 7427
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 11609
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 12605
;   OTHER INFORMATION: unknown
```

; NAME/KEY: unsure
; LOCATION: 12742
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29370
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29422
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29979
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29980
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29981
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30136
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30140
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31205
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31206
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31592
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33095
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33160
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34066
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34072
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 36816
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 39020
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42164
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42459
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46808
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46823
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46826
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 47291
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52786
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52787
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 53384
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 54684
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59215
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59235
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59242
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 63290
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 66614
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68660
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68697
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68718
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68733
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68739
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 69785
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79134
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79198
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 86336
; OTHER INFORMATION: unknown
; OTHER INFORMATION:

US-09-791-211-3

Query Match 2.9%; Score 40; DB 4; Length 87543;
Best Local Similarity 50.8%; Pred. No. 1.1;
Matches 120; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy      269 TTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGT 328
            ||  ||| ||  | |||     | | | | |||||   || | ||  |  | ||||  ||
Db    49995 TTGGTAAAAATAAAATATGTATATATAAATAAAAAAAGAATTTTTAAATCTAAAAAACGT 49936

Qy      329 GGGGCAGAAATATACACAGAAGAGTACTTTAA-CAATGCGCAACCAAGGCAGATTCACAA 387
              | ||| |  | | ||| | |   ||| | ||| | ||   ||  | |   ||||
Db    49935 TTTAAAATAATTTTAAAATAAGTGAATACTAAGCTATGTGGAATACAGAAATAAATACAA 49876

Qy      388 CTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGT 447
            |  ||   ||     |  | | || || |  |  || || | ||| ||  | || ||
Db    49875 TTATGTTAAAAGAGAAATAGAAGGTGAAAAGGAAGAAAAAAATTTAAAGAGAAATGAAGA 49816

Qy      448 GCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAA 503
                | |||| ||| | || | | || || || |  |  |||   | ||  ||
Db    49815 AATAAAAAGAATGCATGAAAAGTGAAAAATATTGAAGCTGGGTAATGAAGATCCAA 49760
```

```
RESULT 9
US-09-397-787-171/c
; Sequence 171, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 171
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-397-787-171

  Query Match             2.8%;  Score 39.4;  DB 4;  Length 435;
  Best Local Similarity   71.2%;  Pred. No. 0.11;
  Matches   52;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        825 TTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCA 884
              ||||||    ||||  ||| ||| |||   |||||||| | |||||||  ||||  |
Db        372 TTCTTCACTGGCGCTTTTTCTTCAGCTTCCTCATCATCAAAATCATCATCATCATCT 313

Qy        885 CCATCATCATCAT 897
              ||||||||||| |
Db        312 TCATCATCATCCT 300


RESULT 10
US-09-328-475C-64/c
; Sequence 64, Application US/09328475C
; Patent No. 6476207
; GENERAL INFORMATION:
;  APPLICANT: Zhang, Jimmy
;  APPLICANT:  Astel, Jon H.
;  APPLICANT:  Carroll III, Eddie
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Ford, Donna M.
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Steinmann, Kathleen E.
;  TITLE OF INVENTION: GENES AND GENE EXPRESSION PRODUCTS THAT
;  TITLE OF INVENTION:  ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER
;  FILE REFERENCE: 1532.002/200130.463
;  CURRENT APPLICATION NUMBER: US/09/328,475C
;  CURRENT FILING DATE:  1999-06-09
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
;   LENGTH: 1024
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1024)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-64

  Query Match             2.8%;  Score 39.4;  DB 4;  Length 1024;
  Best Local Similarity   71.2%;  Pred. No. 0.17;
  Matches   52;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        825 TTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATCATCGCCATCCCCA 884
              ||||||    ||||  ||| ||| |||   |||||||| | |||||||  ||||  |
Db        440 TTCTTCACTGGCGCTTTTTCTTCAGCTTCCTCATCATCAAAATCATCATCATCATCT 381

Qy        885 CCATCATCATCAT 897
              ||||||||||| |
Db        380 TCATCATCATCCT 368


RESULT 11
US-09-058-260-21
; Sequence 21, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 1753
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E013
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-21

  Query Match             2.8%;  Score 39.2;  DB 3;  Length 1753;
  Best Local Similarity   56.1%;  Pred. No. 0.25;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |  |||| ||| |      | ||
Db         24 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 83

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
              |   | |||||  || || |||   |||||||| ||    |||| |  || |||| |
Db         84 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 143

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        144 CGTTGGTGGAAA 155


RESULT 12
US-09-058-260-31
; Sequence 31, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
```

```
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 1756
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E027
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (130)..(1632)
US-09-058-260-31

  Query Match             2.8%;  Score 39.2;  DB 3;  Length 1756;
  Best Local Similarity   56.1%;  Pred. No. 0.26;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |       |  ||
Db         26 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 85

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |   | |||||  || || |||   ||||||||  ||    |||| |  || |||| |
Db         86 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 145

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        146 CGTTGGTGGAAA 157


RESULT 13
US-09-058-260-23
; Sequence 23, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 1776
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E015
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-23

  Query Match             2.8%;  Score 39.2;  DB 3;  Length 1776;
  Best Local Similarity   56.1%;  Pred. No. 0.26;
  Matches   74;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |       |  ||
Db         24 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 83

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                |   | |||||  || || |||   ||||||||  ||    |||| |  || |||| |
Db         84 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 143

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        144 CGTTGGTGGAAA 155


RESULT 14
US-09-058-260-5
; Sequence 5, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1952
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E011
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (197)..(1699)
```

US-09-058-260-5

Query Match 2.8%; Score 39.2; DB 3; Length 1952;
Best Local Similarity 56.1%; Pred. No. 0.27;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |       | ||
Db         93 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 152

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                | |||||  || || |||   |||||||| ||    |||| |  || |||| |
Db        153 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 212

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        213 CGTTGGTGGAAA 224
```

RESULT 15
US-08-781-802-5
; Sequence 5, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,802
; FILING DATE: 10-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/694,078
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/019,580
; FILING DATE: 12-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,704
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,995
; FILING DATE: 01-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 197..1699
; OTHER INFORMATION: /note= "E011 sequence of longest
; OTHER INFORMATION: open reading frame; other possible start codons ATG/met4;
; OTHER INFORMATION: TTG/leu7; GTG/val8; GTG/val15; GTG/val36; ATG/met62"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 197..1699
US-08-781-802-5

Query Match 2.8%; Score 39.2; DB 2; Length 4090;
Best Local Similarity 56.1%; Pred. No. 0.39;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        205 TCAAGTACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCA 264
              ||  || |  |||||||| |  |  |    ||  |   |||| ||| |       | ||
Db         93 TCCGGTGCGATGGATTGCTTCAGGGGAACTTTTAAACACTTGAGTTTGACAACCACTCCT 152

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
                | |||||  || || |||   |||||||| ||    |||| |  || |||| |
Db        153 TAATCATTTAAGATTTAAATGAAAATTAAAATAAATCAAAAAGAGTGATTCAAATGAATA 212

Qy        325 CGGTGGGGCAGA 336
              || ||| | | |
Db        213 CGTTGGTGGAAA 224
```

Search completed: January 21, 2004, 22:19:44
Job time : 98.7682 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:11:45 ; Search time 472.03 Seconds
(without alignments)
10461.606 Million cell updates/sec

Title: US-09-936-869-2
Perfect score: 1401
Sequence: 1 aaactcgacagcaaatatga.....ggaaatttttgtgtacaaatc 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49.6	3.5	22684	13	US-10-017-161-725	Sequence 725, App
2	49.6	3.5	29221	12	US-10-292-798-635	Sequence 635, App
c 3	45	3.2	519	10	US-09-878-574-4292	Sequence 4292, Ap
4	43.8	3.1	625	13	US-10-027-632-44092	Sequence 44092, A
5	43.8	3.1	625	13	US-10-027-632-44093	Sequence 44093, A
6	43.8	3.1	625	13	US-10-027-632-44094	Sequence 44094, A
7	43.8	3.1	625	13	US-10-027-632-44095	Sequence 44095, A
8	43.8	3.1	625	14	US-10-027-632-44092	Sequence 44092, A
9	43.8	3.1	625	14	US-10-027-632-44093	Sequence 44093, A
10	43.8	3.1	625	14	US-10-027-632-44094	Sequence 44094, A
11	43.8	3.1	625	14	US-10-027-632-44095	Sequence 44095, A
c 12	43.6	3.1	26320	13	US-10-034-650-13	Sequence 13, Appl
c 13	43.6	3.1	176373	14	US-10-095-407-17	Sequence 17, Appl
14	43.4	3.1	409	11	US-09-918-995-6789	Sequence 6789, Ap
15	43.4	3.1	13326	13	US-10-311-455-1685	Sequence 1685, Ap
c 16	43.2	3.1	1840	12	US-10-292-798-2003	Sequence 2003, Ap
c 17	43.2	3.1	1840	13	US-10-017-161-2359	Sequence 2359, Ap
c 18	43.2	3.1	18434	13	US-10-311-455-1979	Sequence 1979, Ap
19	43.2	3.1	3673778	13	US-10-312-841-1	Sequence 1, Appli
c 20	43	3.1	201	9	US-09-864-761-18750	Sequence 18750, A
c 21	43	3.1	201	9	US-09-864-761-21744	Sequence 21744, A
c 22	43	3.1	395	9	US-09-864-761-5016	Sequence 5016, Ap
c 23	43	3.1	406	9	US-09-864-761-2003	Sequence 2003, Ap
24	43	3.1	3066	13	US-10-148-687-8	Sequence 8, Appli
c 25	42.8	3.1	428	10	US-09-960-352-573	Sequence 573, App
26	42.6	3.0	491	11	US-09-918-995-11962	Sequence 11962, A
27	42.6	3.0	753	13	US-09-814-353-2034	Sequence 2034, Ap
28	42.6	3.0	753	13	US-09-814-353-8380	Sequence 8380, Ap
29	42.6	3.0	8530	13	US-10-311-455-1727	Sequence 1727, Ap
30	42.4	3.0	143306	10	US-09-729-920-3	Sequence 3, Appli
31	42.2	3.0	327	10	US-09-960-352-4630	Sequence 4630, Ap
c 32	42.2	3.0	7657	13	US-10-311-455-1995	Sequence 1995, Ap
c 33	42.2	3.0	7657	15	US-10-239-676-185	Sequence 185, App
c 34	42.2	3.0	12237	13	US-10-311-455-2331	Sequence 2331, Ap
35	42.2	3.0	16509	13	US-10-311-455-1293	Sequence 1293, Ap
c 36	42	3.0	446	10	US-09-960-352-3400	Sequence 3400, Ap
c 37	42	3.0	1954	9	US-09-925-301-82	Sequence 82, Appl
c 38	42	3.0	2812	15	US-10-198-846-9693	Sequence 9693, Ap
39	42	3.0	152331	14	US-10-095-407-16	Sequence 16, Appl
40	41.8	3.0	642	11	US-09-866-050A-532	Sequence 532, App
41	41.8	3.0	642	15	US-10-152-661-532	Sequence 532, App
c 42	41.8	3.0	11577	13	US-10-311-455-414	Sequence 414, App
43	41.6	3.0	172	9	US-09-864-761-25620	Sequence 25620, A
44	41.6	3.0	511	9	US-09-864-761-8959	Sequence 8959, Ap
c 45	41.6	3.0	6071	13	US-10-311-455-340	Sequence 340, App

ALIGNMENTS

RESULT 1
US-10-017-161-725
; Sequence 725, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 725
; LENGTH: 22684
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(22684)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(615)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (753)..(878)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17062)..(17263)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (17606)..(17723)
; FEATURE:
; NAME/KEY: CDS

```
;   LOCATION: (19946)..(20099)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20577)..(20637)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22148)..(22484)
US-10-017-161-725

  Query Match             3.5%;  Score 49.6;  DB 13;  Length 22684;
  Best Local Similarity   61.7%;  Pred. No. 0.074;
  Matches   79;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        812 CCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATC 871
              ||  |  ||| ||||| || ||       || | ||  |   ||||| ||||| | ||||
Db      10903 CCTCCTCCTCCTCTTCATCCTCCTTCTTCTTCTATCATCATCATCATTATCATCATCATC 10962

Qy        872 ATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTG 931
              |||  ||||  || |||||||||||| |  |  ||||  ||| || |      | || ||
Db      10963 ATCATCATCATCATCATCATCATCACAGGGTCTCACTCTGTCACCCAGGCTGGAGTG 11022

Qy        932 TAGAGGAA 939
              || || |
Db      11023 CAGTGGCA 11030


RESULT 2
US-10-292-798-635
; Sequence 635, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 635
;   LENGTH: 29221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(29221)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(303)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (441)..(566)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15168)..(15253)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (16842)..(16951)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (17294)..(17411)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19634)..(19787)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20265)..(20325)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (21836)..(21908)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (26469)..(26554)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (26766)..(26901)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (27401)..(27442)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (28881)..(29021)
US-10-292-798-635

  Query Match             3.5%;  Score 49.6;  DB 12;  Length 29221;
  Best Local Similarity   61.7%;  Pred. No. 0.086;
  Matches   79;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        812 CCGACCGCTCATCTTCTTCTTCTGCGCAATTTTTTCTGCTGGATCATCATCATTACCATC 871
              ||  |  ||| ||||| || ||       || | ||  |   ||||| ||||| | ||||
Db      10591 CCTCCTCCTCCTCTTCATCCTCCTTCTTCTTCTATCATCATCATCATTATCATCATCATC 10650

Qy        872 ATCGCCATCCCCACCATCATCATCATGATGGTATCTCTATCTCTCCCTGGCAATCGATTG 931
              |||  ||||  || |||||||||||| |  |  ||||  ||| || |      | || ||
Db      10651 ATCATCATCATCATCATCATCATCACAGGGTCTCACTCTGTCACCCAGGCTGGAGTG 10710

Qy        932 TAGAGGAA 939
              || || |
Db      10711 CAGTGGCA 10718


RESULT 3
US-09-878-574-4292/c
; Sequence 4292, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 4292
;   LENGTH: 519
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(519)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3028-013-Q1-B1-D1
US-09-878-574-4292

  Query Match             3.2%;  Score 45;  DB 10;  Length 519;
  Best Local Similarity   47.7%;  Pred. No. 0.14;
  Matches  132;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        265 ACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
              | |||||| |  ||| ||| || |||||||||||   ||||  ||  |  ||   |||||
Db        511 AAAATTTAAAGTAAAAATAAAATATAAAAAATATGTAAATAAAAAAAATAAGATAAGAAAA 452

Qy        325 CGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCA 384
              | ||       ||  | | |     |     | |  || ||       |   |||  | | | |
Db        451 AAGAGAAAACAAATAAAAAAGTATAATCAATTAAAAAAAAAACAAGTAAAGATAAAATAA 392
```

```
Qy        385 CAACTTGATTTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGG 444
              ||  | |          |  || |  | ||||   ||   | ||| || |  | | ||
Db        391 AAAAGTTAAAAAATATAATATAAAAAAAAATAAATTTGAATAAAAAAAATGATAAAATGA 332

Qy        445 AGTGCATTTGAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAAT 504
              | |  |    || ||  ||  | |   | || |   |  | |||| ||  | ||  |
Db        331 ATTTAAGAAAAATATATATATAAAATAAAAAATAATAATAATAAATATATATATTAAGAA 272

Qy        505 GCAAGAGTGCATTTCCCTATTATTTCCAGAAATGTAT 541
                ||  |  | |      | ||||   |  ||  |||
Db        271 TAAAAATTATAATGGATAAATATTGAAAAGAAAATAT 235


RESULT 4
US-10-027-632-44092
; Sequence 44092, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44092
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44092

  Query Match             3.1%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| || ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |   | | | ||  |  |||||||  | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 5
US-10-027-632-44093
; Sequence 44093, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44093
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44093

  Query Match             3.1%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| || ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |   | | | ||  |  |||||||  | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 6
US-10-027-632-44094
; Sequence 44094, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
```

```
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44094
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44094

  Query Match             3.1%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| || ||  |||    |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |    | | | ||   | |||||| | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 7
US-10-027-632-44095
; Sequence 44095, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44095
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44095

  Query Match             3.1%;  Score 43.8;  DB 13;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| || ||  |||    |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |    | | | ||   | |||||| | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 8
US-10-027-632-44092
; Sequence 44092, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44092
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44092

  Query Match             3.1%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||| || ||  |||    |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              || | |||| | | |    | | | ||   | |||||| | | |||||  |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 9
US-10-027-632-44093
; Sequence 44093, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
```

```
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44093
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44093

  Query Match             3.1%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | | |    | | | ||   |  |||||||  | |  |||||   |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 10
US-10-027-632-44094
; Sequence 44094, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44094
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44094

  Query Match             3.1%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | | |    | | | ||   |  |||||||  | |  |||||   |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
              ||||    | | |||
Db        469 GCAGTCAAATAGCTT 483


RESULT 11
US-10-027-632-44095
; Sequence 44095, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44095
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(625)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-44095

  Query Match             3.1%;  Score 43.8;  DB 14;  Length 625;
  Best Local Similarity   57.8%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        256 GAGAAGCCAACAATTTATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGG 315
              ||||       |  | | |||  || | ||||||  | |||||  ||  ||  |||   |
Db        349 GAGAGAGACCCTGTCTCTAAATAAAAAATAAAAATATAAATACAATAAAATAAAAATAAA 408

Qy        316 AAGTGAAAACGGTGGGGCAGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAG 375
              ||  | ||||  | | |    | | | ||   |  |||||||  | |  |||||   |  ||| ||
Db        409 AAATAAAAAGGATAGACAAAATAGATCAATTGAAGAGTCCATAAACAAACCCAAACAAAT 468

Qy        376 GCAGATTCACAACTT 390
```

```
Db        469 GCAGTCAAATAGCTT 483
```

RESULT 12
US-10-034-650-13/c
; Sequence 13, Application US/10034650
; Publication No. US20030216558A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000128
; CURRENT APPLICATION NUMBER: US/10/034,650
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 09/474,377
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 26320
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-034-650-13

Query Match 3.1%; Score 43.6; DB 13; Length 26320;
Best Local Similarity 49.6%; Pred. No. 3.4;
Matches 112; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```
Qy        277 AAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGA 336
Db       6178 AGAAGTAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAAGAGAGAGGGGGGGAGAGA 6119

Qy        337 AATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTC 396
Db       6118 AAGAGAGAGAAAGAAAGAGAGAGAGAGGGGGAGAGAAAGAGAGAGAGAAAGAGAAAGA 6059

Qy        397 TGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTGAA 456
Db       6058 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGGAAGAAAGA 5999

Qy        457 AATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATA 502
Db       5998 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA 5953
```

RESULT 13
US-10-095-407-17/c
; Sequence 17, Application US/10095407
; Publication No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-17

Query Match 3.1%; Score 43.6; DB 14; Length 176373;
Best Local Similarity 48.8%; Pred. No. 10;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```
Qy        274 AAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGC 333
Db     122513 AAGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAAGAGAAGAAA 122454

Qy        334 AGAAATATACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGAT 393
Db     122453 GAAAGTAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 122394

Qy        394 TTCTGGACCTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTT 453
Db     122393 AGAAAGAGAAAGAAAGGGAGAGAAAGAAGGAAGGAAGGAAAGAAAGAAAGCAAGCAAGCA 122334

Qy        454 GAAAATGAATGGAGAGCGCACAAAATGGAGGACGAATAAATGAAATATAATGCAAGAGTG 513
Db     122333 GGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGATAAAGAAAGAAAGAAAGAAAG 122274

Qy        514 CA 515
Db     122273 AA 122272
```

RESULT 14
US-09-918-995-6789
; Sequence 6789, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6789
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(409)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-6789

Query Match 3.1%; Score 43.4; DB 11; Length 409;
Best Local Similarity 52.9%; Pred. No. 0.32;
Matches 92; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy        151 GCTTGGCTTCTGGGTAATATGAGAAGTAAAGAAGTAAAAGGTATTTGACTCTAGTCAAGT 210
Db        164 GTTTGGGTTTGGGTTAGAAGGAGAATGAAATAAGAGAGATTGTATTTAAACTAGCCTAGG 223

Qy        211 ACATTGGATTGCCTTTGTCGGGGCTTGGATGGCTTGGGTTCGTGTGAGAAGCCAACAATT 270
Db        224 AAAATGGTGCACTGAAATAAAGGTTACTGCAGGGGTTTTCCCCATGACAAGCANAAAAAA 283

Qy        271 TATAAGAAATATATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAA 324
Db        284 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAGAAAAA 337
```

RESULT 15
US-10-311-455-1685
; Sequence 1685, Application US/10311455

; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1685
; LENGTH: 13326
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1685

Query Match 3.1%; Score 43.4; DB 13; Length 13326;
Best Local Similarity 53.2%; Pred. No. 2.6;
Matches 92; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy        282 TATAAAATAAAAAATAAAAAAATTTAAGTGTTGGAAGTGAAAACGGTGGGGCAGAAATAT 341
              | | |||  ||  | ||| |||||  |   ||| | |  | |  | | |||    ||| |
Db       9237 TTTTAAAATAAGGAGAAATAAATTATATGTTTGTATGGTAGAGGGATAGGGAGAGAATTT 9296

Qy        342 ACACAGAAGAGTACTTTAACAATGCGCAACCAAGGCAGATTCACAACTTGATTTCTGGAC 401
              | |||| || | ||||   |||    |  | |  | | | |  | |  |||    |
Db       9297 TTAGAGAATAGAAATTTAGGGATGAAGTAGGAGGTGATAGTTATTAATAAATTAAGTGTT 9356

Qy        402 CTCGAATACGAGATAATGGTGGTAAGAAATAAAGGAAGAGTGGAGTGCATTTG 454
              | ||| |  ||| ||||  || | || ||   | |||  |   |||  ||||
Db       9357 ATAGAAGAAAAGAAAATGTAGGGATGAGATTGGGAAAGTATATTGTGGTTTTG 9409
```

Search completed: January 22, 2004, 01:20:35
Job time : 475.03 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:02:36 ; Search time 1813.1 Seconds
(without alignments)
12185.050 Million cell updates/sec

Title: US-09-936-869-3
Perfect score: 909
Sequence: 1 atgggtcgtgggaagattga.....cgcatgttatctttctgttg 909

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	310.2	34.1	697	14	CD482321	CD482321 atr01-19m
2	278.2	30.6	823	14	CB973770	CB973770 CAB30003_
3	276.8	30.5	819	14	CB975703	CB975703 CAB30007_
4	275.4	30.3	668	14	CB009226	CB009226 VVC065C11
5	274.6	30.2	665	14	CB005658	CB005658 VVC024D03
6	273.8	30.1	786	14	CB969491	CB969491 CAB10002_
7	273.8	30.1	843	14	CB969483	CB969483 CAB10002_
8	271	29.8	754	12	BI311053	BI311053 EST531280
9	267	29.4	608	9	AW278878	AW278878 sf99h09.y
10	264.2	29.1	822	14	CB974938	CB974938 CAB30005_
11	262.6	28.9	824	14	CB975572	CB975572 CAB30007_
12	260	28.6	858	14	CA857928	CA857928 EST635183
13	258.6	28.4	615	13	BU045147	BU045147 PP_LEa002
14	258.6	28.4	631	13	BU043610	BU043610 PP_LEa001
15	257.8	28.4	649	9	AW184799	AW184799 se82f12.y
16	257.2	28.3	610	13	BU046256	BU046256 PP_LEa002
17	256.6	28.2	866	10	BG441292	BG441292 GA_Ea001
18	255.2	28.1	634	14	CA902463	CA902463 PCS02695F
19	254.6	28.0	715	12	BI933168	BI933168 EST553057
20	252.8	27.8	892	10	BG444639	BG444639 GA_Ea002
21	251.6	27.7	605	13	BU045737	BU045737 PP_LEa002
22	251.4	27.7	761	9	AI731368	AI731368 BNLGHi937
23	250.8	27.6	776	10	BG445079	BG445079 GA_Ea002
24	249.8	27.5	754	9	AI731375	AI731375 BNLGHi938
25	247.6	27.2	603	14	CB911729	CB911729 VVD135A10
26	247.6	27.2	605	14	CB917153	CB917153 VVD015E04
27	246.2	27.1	548	10	BE434089	BE434089 EST405167
28	246	27.1	538	10	BE433121	BE433121 EST399650
29	245.2	27.0	607	9	AW705451	AW705451 sk49c05.y
30	245	27.0	624	14	CB921120	CB921120 VVD050C09
31	245	27.0	663	14	CB918484	CB918484 VVD033H03
32	244.4	26.9	609	13	BU042190	BU042190 PP_LEa001
33	243.2	26.8	707	9	AI898432	AI898432 EST267875
34	241.4	26.6	665	9	AI487420	AI487420 EST245742
35	241.4	26.6	697	14	CB290594	CB290594 UCRCS01_0
36	241.2	26.5	632	13	BU048762	BU048762 PP_LEa003
37	241.2	26.5	686	13	BQ411600	BQ411600 GA_Ed004
38	239.6	26.4	656	13	BQ411583	BQ411583 GA_Ed004
39	239.6	26.4	698	10	BE659915	BE659915 7-H4 Gmax
40	237.2	26.1	537	9	AI486290	AI486290 EST244611
41	237.2	26.1	539	9	AI771796	AI771796 EST252896
42	236.4	26.0	618	9	AI484747	AI484747 EST243008
43	236.4	26.0	729	14	CB066648	CB066648 EST645271
44	236.4	26.0	748	12	BI311095	BI311095 EST531284
45	235	25.9	624	9	AW277843	AW277843 sf87d01.y

ALIGNMENTS

RESULT 1
CD482321
LOCUS CD482321 697 bp mRNA linear EST 04-JUN-2003
DEFINITION atr01-19ms1-a12 Atr01 Amborella trichopoda cDNA clone atr01-19ms1-a12 5', mRNA sequence.
ACCESSION CD482321
VERSION CD482321.1 GI:31403589
KEYWORDS EST.
SOURCE Amborella trichopoda
ORGANISM Amborella trichopoda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; stem Magnoliophyta; Amborellaceae; Amborella.
REFERENCE 1 (bases 1 to 697)
AUTHORS dePamphilis,C., Soltis,D., Soltis,P., Farmerie,W., Ma,H., Tanksley,S., Leebens-Mack,J., Field,D., Buzgo,M., Kim,S., Ilut,D., Landherr,L., Hu,Y., Wall,K., Albert,V., Carlson,J., Doyle,J., Frohlich,M., Miller,W., Oppenheimer,D. and Theissen,G.
TITLE Generation of ESTs from early flower buds of Amborella trichopoda
JOURNAL Unpublished
COMMENT Contact: Claude dePamphilis or James Leebens-Mack
Mueller Laboratory
Penn State University
208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn State University, University Park, PA 16802, USA
Tel: 814 863 6413
Fax: 814 865 9131

```
            Email: cwd3@psu.edu or jhl10@psu.edu
            The sequence provided is trimmed of vector and low quality regions.
            Full sequence and original trace file are available from the Plant
            Genome Network website (http://pgn.cornell.edu)
            Plate: atr01-19ms1  row: a  column: 12
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Amborella trichopoda"
                     /mol_type="mRNA"
                     /db_xref="taxon:13333"
                     /clone="atr01-19ms1-a12"
                     /tissue_type="flower buds"
                     /dev_stage="<= 2.5mm buds"
                     /lab_host="SOLR"
                     /clone_lib="Atr01"
                     /note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
                     Site_2: Xho1; Amborella trichopoda Baill; This library was
                     made from male flowers only. Only floral buds with
                     diameter of 2.5 mm of less were used for RNA isolation.
                     This is a directionally cloned, non-normalized library.
                     Avg. insert length: 1611; Primers: M13F and M13R;
                     Antibiotic: 50 ug/ml Ampicillin; Primary Titer: 2.24E6 pfu
                     total; Amplified Titer: 1.37E10 pfu/ml; Mass Excised
                     Titer: 3.53E9 total; This library has been generated by
                     the Floral Genome Project (FGP). We would like to thank
                     David Lorence at the National Tropical Botanical Garden
                     for providing plant material for library building. The
                     Floral Genome Project is funded by NSF's Plant Genome
                     Research Program (DBI-0115684). More information about the
                     project can be obtained at http://fgp.bio.psu.edu"
BASE COUNT      244 a    138 c    151 g    163 t      1 others
ORIGIN

  Query Match             34.1%;  Score 310.2;  DB 14;  Length 697;
  Best Local Similarity   68.3%;  Pred. No. 1.2e-74;
  Matches  445;  Conservative    0;  Mismatches  204;  Indels    3;  Gaps    1;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || ||||||||||| ||| | || || || || || ||  | ||||| ||
Db         46 ATGGGAAGGGGAAAGATTGAGATCAAGCGCATAGAGAACACAACTAATAGGCAGGTGACC 105

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              || ||||| || || ||||| || | ||||| || |||||||||||| ||||| ||||||
Db        106 TTTTGCAAACGGCGCAATGGGCTACTCAAGAAAGCCTATGAATTATCTGTTCTCTGTGAT 165

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || ||||| || || ||||| || || |||| |||| || |||| ||| ||  ||
Db        166 GCTGAAGTTGCACTTATTGTCTTTTCAAGTCGAGGTCGACTGTACGAATATGCAAATAAC 225

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              |||||||| |  || ||||| |||||||||||| | |||| ||| | |  |||  ||
Db        226 AGCGTGAAAACAACAATTGATAGGTACAAGAAGGCATGCGCTGATAGCTCTCACTCCGGA 285

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              |  | |||||| |||||||||| |||| ||||| || |||| |||||  |  | ||
Db        286 ACTGTTTCAGAGGCCAATTCTCAATATTACCAACAAGAAGCTGCAAAACTTCGCAATCAA 345

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||  |  ||||     ||| |||| ||||| ||||||||||||    ||    |  |||
Db        346 ATCCAGGTTTTAACGAATACAAACAGGCAGTTGATGGGTGACTCCGTTGGTTCAATGACT 405

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||  |||| |  | | |||     | | ||||| ||||| ||   | |  ||||
Db        406 GTTAAGGAGCTCAGGACATTGGAGAACAAATTGGAAAAGGGAATTAGTAAAATTAGATCA 465

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              |||||||| ||| |  |  | |  || ||||||  ||||||||    ||||||    |||
Db        466 AAAAAGAATGAGCTACTATTCGCTGAAATCGACTACATGCAGAATCGGGAACTTGAACTA 525

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              |  || ||   | |||| ||    ||||||||| ||   | | |     ||  ||| |
Db        526 CAGAAAGACAACATGCTTCTCCGAGCTAAGATAGCTGAAAATGAAAGGGCACAACACATG 585

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              ||||||||  |||  || |||||||||| | ||||  |||| |||||||| || | |||
Db        586 AACATGTTGCCAGGGCCAGAATATGATGTATTGCCANCATTTGACTCTCGGAACTACCTC 645

Qy        601 CATGCAAATCTAATCGATGCGGCCCATC---ACTATGCACATCAGGAACAAA 649
              || |  || ||| | ||  ||  |||||   ||||  | |||||||| ||||
Db        646 CAAGTTAACCTATTAGAACCGAACCATCATAACTACTCGCATCAGGAGCAAA 697


RESULT 2
CB973770
LOCUS       CB973770                823 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB30003_IVa_Fa_G05 Cabernet Sauvignon Berry Stage I - CAB3 Vitis
            vinifera cDNA clone CAB30003_IVa_Fa_G05 5', mRNA sequence.
ACCESSION   CB973770
VERSION     CB973770.1  GI:30296976
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .823
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB30003_IVa_Fa_G05"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry stage I"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry Stage I - CAB3"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB3 is a cDNA library of Vitis vinifera 'Cabernet
                     Sauvignon' Clone 8 berries. Samples were collected after
                     berry set from field-grown vines during stage I of berry
                     growth, 17 days after full bloom. The average berry size
                     was 6 milimeters. Sampled vines were located at the
                     University of California, Davis, Experimental Vineyard.
                     cDNAs were made by oligo-dT priming and directionally
                     cloned. 5'and 3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      256 a    174 c    199 g    193 t      1 others
ORIGIN

  Query Match             30.6%;  Score 278.2;  DB 14;  Length 823;
  Best Local Similarity   65.6%;  Pred. No. 9.4e-66;
  Matches  406;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || ||||| ||||| |||||||| ||||| || || ||||| |||||||||
Db         89 ATGGGGAGAGGAAAGATCGAGATCAAGAGGATCGAAAACACGACCAACCGTCAGGTCACA 148

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
```

```
Db        149 TTCTGCAAGCGAAGGAATGGGCTTTTGAAGAAGGCTTATGAATTATCAGTGCTATGTGAT 208

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        209 GCAGAAGTTGCCCTCATCGTCTTCTCCAGCCGCGGTCGAGTCTATGAGTACTCAAACAAC 268

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        269 AACATAAAATCAACCATAGATAGGTACAAGAAGGCCAGCTCAGATAGTACAAATGGAGGC 328

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        329 TCTACCATGGAGATCAATGCCCAATATTACCAGCAAGAATCAGCAAAGCTGCGCCAGCAA 388

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        389 ATACAGATGCTGCAGAATTCTAACAGGCACTTAATGGGTGATTCCTTGGCTTCCTTGACT 448

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        449 GTGAAGGAGCTAAAGCAGCTCGAGAACAGGCTTGTACGAGGCATCACAAGAATCAGGTCG 508

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        509 AAGAAGCATGAGTTGCTGTTGGCTGAGATTGAATACTTGCAGAAAAGGGAAATTGAGCTG 568

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        569 GAAAATGAAAGCGTATATCTCCGAACCAAGATTGCAGAAGTGGAGAGGCTTCAGCAAGCA 628

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        629 AACATGGTATCAACACATGAGTTCAATGCCATCCAGGCATTAGTTTCTCGCAATTTCTTT 688

Qy        601 CATGCAAATCTAATCGATG 619

Db        689 CAGCCCAATATGATTGAGG 707
```

RESULT 3
CB975703

```
LOCUS       CB975703                819 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB30007_Ib_Fb_D07 Cabernet Sauvignon Berry Stage I - CAB3 Vitis
            vinifera cDNA clone CAB30007_Ib_Fb_D07 5', mRNA sequence.
ACCESSION   CB975703
VERSION     CB975703.1  GI:30298909
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 819)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .819
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB30007_Ib_Fb_D07"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry stage I"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry Stage I - CAB3"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB3 is a cDNA library of Vitis vinifera 'Cabernet
                     Sauvignon' Clone 8 berries. Samples were collected after
                     berry set from field-grown vines during stage I of berry
                     growth, 17 days after full bloom. The average berry size
                     was 6 milimeters. Sampled vines were located at the
                     University of California, Davis, Experimental Vineyard.
                     cDNAs were made by oligo-dT priming and directionally
                     cloned. 5'and 3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      262 a    171 c    194 g    191 t      1 others
ORIGIN
```

Query Match 30.5%; Score 276.8; DB 14; Length 819;
Best Local Similarity 65.6%; Pred. No. 2.3e-65;
Matches 404; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

```
Qy          4 GGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTTTC 63

Db         72 GGGAGAGGAAAGATCGAGATCAAGAGGATCGAAAACACGACCAACCGTCAGGTCACATTC 131

Qy         64 TGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATGCA 123

Db        132 TGCAAGCGAAGGAATGGGCTTTTGAAGAAGGCTTATGAATTATCAGTGCTATGTGATGCA 191

Qy        124 GAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACAGC 183

Db        192 GAAGTTGCCCTCATCGTCTTCTCCAGCCGCGGTCGAGTCTATGAGTACTCAAACAACAAC 251

Qy        184 GTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGGCG 243

Db        252 ATAAAATCAACCATAGATAGGTACAAGAAGGCCAGCTCAGATAGTACAAATGGAGGCTCT 311

Qy        244 ATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGATT 303

Db        312 ACCATGGAGATCAATGCCCAATATTACCAGCAAGAATCAGCAAAGCTGCGCCAGCAAATA 371

Qy        304 GACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACATT 363

Db        372 CAGATGCTGCAGAATTCTAACAGGCACTTAATGGGTGATTCCTTGGCTTCCTTGACTGTG 431

Qy        364 AAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCAAA 423

Db        432 AAGGAGCTAAAGCAGCTCGAGAACAGGCTTGAACGAGGCATCACAAGAATCAGGTCGAAG 491

Qy        424 AAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTATC 483

Db        492 AAGCATGAGTTGCTGTTGGCTGAGATTGAATACTTGCAGAAAAGGGAAATTGAGCTGGAA 551

Qy        484 CAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACGAAC 543

Db        552 AATGAAAGCGTATATCTCCGAACCAAGATTGCAGAAGTGGAGAGGCTTCAGCAAGCAAAC 611

Qy        544 ATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTACAT 603

Db        612 ATGGTATCAACACATGAGTTCAATGCCATCCAGGCATTAGTTTCTCGCAATTTCTTTCAG 671

Qy        604 GCAAATCTAATCGATG 619

Db        672 CCCAATATGATTGAGG 687
```

RESULT 4
CB009226

```
LOCUS       CB009226                668 bp    mRNA    linear   EST 10-JAN-2003
```

```
DEFINITION  VVC065C11_230854 An expressed sequence tag database for abiotic
            stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
            cDNA clone VVC065C11 5, mRNA sequence.
ACCESSION   CB009226
VERSION     CB009226.1  GI:27586531
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Cushman,J.C.
  TITLE     An expressed sequence tag database for abiotic stressed berries of
            Vitis vinifera var. Chardonnay
  JOURNAL   Unpublished
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 065  row: C  column: 11
            Seq primer: T3 20mer
            High quality sequence stop: 668.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /db_xref="taxon:29760"
                     /clone="VVC065C11"
                     /tissue_type="berries"
                     /dev_stage="mixed; 8, 9, 11, 13, 15, 16 weeks daf"
                     /clone_lib="An expressed sequence tag database for abiotic
                     stressed berries of Vitis vinifera var. Chardonnay"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      221 a    137 c    166 g    144 t
ORIGIN
```

```
  Query Match             30.3%;  Score 275.4;  DB 14;  Length 668;
  Best Local Similarity   65.6%;  Pred. No. 5.1e-65;
  Matches  402;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || ||||| ||||| |||||||| ||||| || || ||||| ||||||||
Db         56 ATGGGGAGAGGAAAGATCGAGATCAAGAGGATCGAAAACACGACCAACCGTCAGGTCACA 115

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||||||||  | |||||  | || |||||||| ||||||||||||||| || ||||||
Db        116 TTCTGCAAGCGAAGGAATGGGCTTTTGAAGAAGGCTTATGAATTATCAGTGCTATGTGAT 175

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              |||||||| |||||||||||||||||||| | ||  || | |||||  |   | ||| ||
Db        176 GCAGAAGTTGCCCTCATCGTCTTCTCCAGCCGCGGTCGAGTCTATGAGTACTCAAACAAC 235

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              | | | ||    || || || |||||||||||| | ||  || | |   | |||||
Db        236 AACATAAAATCAACCATAGATAGGTACAAGAAGGCCAGCTCAGATAGTACAAATGGAGGC 295

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | |     |||  |||| | || || |  || || ||  | |  || || |  | || ||
Db        296 TCTACCATGGAGATCAATGCCCAATACTACCAGCAAGAATCAGCAAAGCTGCGCCAGCAA 355

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||  | ||  |||| ||| | || ||||| || ||||||||      |  |  | ||||
Db        356 ATACAGATGCTGCAGAATTCTAACAGGCACTTAATGGGTGATTCCTTGGCTTCCTTGACT 415

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
```

```
              | ||||| || ||||| || |||    | ||||||  ||| ||||  || |  | ||
Db        416 GTGAAGGAGCTAAAGCAGCTCGAGAACAGGCTTGAACGAGGCATCACAAGAATCAGGTCG 475

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| | ||| || || | |  ||||| ||    | |||||| |||||||      ||
Db        476 AAGAAGCATGAGTTGCTGTTGGCTGAGATTGAGTACTTGCAGAAAAGGGAAATTGAGCTG 535

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
              |  || |  |    ||| || | |||||| || ||     |||      ||  |  |
Db        536 GAAAATGAAAGCGTATATCTCCGAACCAAGATTGCAGAAGTGGAGAGGCTTCAGCAAGCA 595

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              |||||| ||||| | |  || | ||||  | |  ||||| |  ||||| |||||| |
Db        596 AACATGGTATCAACACATGAGTTCAATGCCATCCAGGCATTAGTTTCTCGCAATTTCTTT 655

Qy        601 CATGCAAATCTAA 613
              ||  | ||| | |
Db        656 CAGCCCAATATGA 668
```

```
RESULT 5
CB005658
LOCUS       CB005658                 665 bp    mRNA    linear   EST 10-JAN-2003
DEFINITION  VVC024D03_137486 An expressed sequence tag database for abiotic
            stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
            cDNA clone VVC024D03 5, mRNA sequence.
ACCESSION   CB005658
VERSION     CB005658.1  GI:27582963
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   Cushman,J.C.
  TITLE     An expressed sequence tag database for abiotic stressed berries of
            Vitis vinifera var. Chardonnay
  JOURNAL   Unpublished
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 024  row: D  column: 03
            Seq primer: T3 20mer
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .665
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /db_xref="taxon:29760"
                     /clone="VVC024D03"
                     /tissue_type="berries"
                     /dev_stage="mixed; 8, 9, 11, 13, 15, 16 weeks daf"
                     /clone_lib="An expressed sequence tag database for abiotic
                     stressed berries of Vitis vinifera var. Chardonnay"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      220 a    137 c    165 g    143 t
ORIGIN
```

```
  Query Match             30.2%;  Score 274.6;  DB 14;  Length 665;
  Best Local Similarity   65.7%;  Pred. No. 8.5e-65;
  Matches  400;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || ||||| ||||| |||||||| ||||| || || ||||| ||||||||
```

```
Db         56 ATGGGGAGAGGAAAGATCGAGATCAAGAGGATCGAAAACACGACCAACCGTCAGGTCACA 115

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db        116 TTCTGCAAGCGAAGGAATGGGCTTTTGAAGAAGGCTTATGAATTATCAGTGCTATGTGAT 175

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        176 GCAGAAGTTGCCCTCATCGTCTTCTCCAGCCGCGGTCGAGTCTATGAGTACTCAAACAAC 235

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        236 AACATAAAATCAACCATAGATAGGTACAAGAAGGCCAGCTCAGATAGTACAAATGGAGGC 295

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        296 TCTACCATGGAGATCAATGCCCAATACTACCAGCAAGAATCAGCAAAGCTGCGCCAGCAA 355

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        356 ATACAGATGCTGCAGAATTCTAACAGGCACTTAATGGGTGATTCCTTGGCTTCCTTGACT 415

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        416 GTGAAGGAGCTAAAGCAGCTCGAGAACAGGCTTGAACGAGGCATCACAAGAATCAGGTCG 475

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        476 AAGAAGCATGAGTTGCTGTTGGCTGAGATTGAGTACTTGCAGAAAAGGGAAATTGAGCTG 535

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540

Db        536 GAAAATGAAAGCGTATATCTCCGAACCAAGATTGCAGAAGTGGAGAGGCTTCAGCAAGCA 595

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600

Db        596 AACATGGTATCAACACATGAGTTCAATGCCATCCAGGCATTAGTTTCTCGCAATTTCTTT 655

Qy        601 CATGCAAAT 609

Db        656 CAGCCCAAT 664


RESULT 6
CB969491
LOCUS       CB969491                786 bp    mRNA    linear   EST 30-APR-2003
DEFINITION  CAB10002_IIIa_Fa_E09 Cabernet Sauvignon Flower Pre-bloom - CAB1
            Vitis vinifera cDNA clone CAB10002_IIIa_Fa_E09 5', mRNA sequence.
ACCESSION   CB969491
VERSION     CB969491.1  GI:30251758
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB10002_IIIa_Fa_E09"
                     /sex="Hermaphrodite"
                     /dev_stage="Pre-bloom"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Flower Pre-bloom - CAB1"
                     /note="Organ: Flower - Pre-bloom; Vector: pDNR; Site_1:
                     SfiI; Site_2: SfiI; CAB1 is a cDNA library of Vitis
                     vinifera cv. 'Cabernet Sauvignon' Clone 8 flowers. Samples
                     were collected approximately eleven days before onset of
                     bloom (clusters at this stage were fully developed and
                     flowers with calyptras or caps still attached. Sampled
                     vines were located at the University of California, Davis,
                     Experimental Vineyard. cDNAs were made by oligo-dT priming
                     and directionally cloned. 5'and 3' adaptors were used in
                     cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      243 a    178 c    181 g    184 t
ORIGIN

  Query Match           30.1%;  Score 273.8;  DB 14;  Length 786;
  Best Local Similarity   64.9%;  Pred. No. 1.5e-64;
  Matches  442;  Conservative    0;  Mismatches  227;  Indels   12;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60

Db         16 ATGGGAAGGGGGAAGATCGAGATCAAGCGGATCGAAAACACCACTAACCGGCAGGTTACT 75

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120

Db         76 TTCTGCAAGCGCCGCAACGGCTTACTCAAGAAGGCCTATGAATTATCTGTCCTATGTGAT 135

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180

Db        136 GCAGAGGTAGCCCTCATCGTCTTCTCCAGCCGCGGCCGCCTCTACGAATATGCCAACAAC 195

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240

Db        196 AGTGTTAAATCAACAATTGAGAGGTATAAGAAGGCAAGCGCTGATTCCTCCAACACCGGG 255

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300

Db        256 TCCGTTTCTGAAGCCAATGCTCAGTTCTACCAGCAAGAATCCTCCAAACTGCATCAGCAG 315

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360

Db        316 ATTCGTAATTTGCAGAATTCGAACAGGCACATGCTGGGTGAGTCTCTGGGCTCCTTGAAT 375

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420

Db        376 TTCAAGGACCTCAAGAGCCTGGAGATTTGGTTAGAGAAAGGCATAAGTAGAATCAGATCC 435

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480

Db        436 AAAAAGAATGAGCTGTTGTTTGCTGAGATCGAGTATATGCAAAAGAGGGAAATAGACTTG 495

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGC---CACAAC 537

Db        496 CATAATGATAACCAGTATCTCCGAGCAAGGATAGCCGAGAATGAGAGAAACGAACAGCAG 555

Qy        538 ACGAACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCA---------TTCGACTCT 588

Db        556 ATGAGCTTGATGCCAGGAGGAGCTAACTATGAACTCATGCCATCTCAGCAGTTCGACTCT 615

Qy        589 CGAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAA 648

Db        616 AGAAACTATTTCCAACTAAATGGACTGCAACCAAATCAATCATACTCTCGCCAAGACCAA 675
```

```
Qy        649 ACAACGCTTCAGCTTGGCTGA 669
              || | |||||  | |  | ||
Db        676 CCAGCTCTTCAATTAGTTTAA 696


RESULT 7
CB969483
LOCUS       CB969483                 843 bp    mRNA    linear   EST 30-APR-2003
DEFINITION  CAB10002_IIIa_Fa_D11 Cabernet Sauvignon Flower Pre-bloom - CAB1
            Vitis vinifera cDNA clone CAB10002_IIIa_Fa_D11 5', mRNA sequence.
ACCESSION   CB969483
VERSION     CB969483.1  GI:30251750
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB10002_IIIa_Fa_D11"
                     /sex="Hermaphrodite"
                     /dev_stage="Pre-bloom"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Flower Pre-bloom - CAB1"
                     /note="Organ: Flower - Pre-bloom; Vector: pDNR; Site_1:
                     SfiI; Site_2: SfiI; CAB1 is a cDNA library of Vitis
                     vinifera cv. 'Cabernet Sauvignon' Clone 8 flowers. Samples
                     were collected approximately eleven days before onset of
                     bloom (clusters at this stage were fully developed and
                     flowers with calyptras or caps still attached. Sampled
                     vines were located at the University of California, Davis,
                     Experimental Vineyard. cDNAs were made by oligo-dT priming
                     and directionally cloned. 5'and 3' adaptors were used in
                     cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      265 a    188 c    191 g    197 t      2 others
ORIGIN

  Query Match           30.1%;  Score 273.8;  DB 14;  Length 843;
  Best Local Similarity   64.9%;  Pred. No. 1.6e-64;
  Matches  442;  Conservative    0;  Mismatches  227;  Indels   12;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | |||||||| ||||| ||| |||| ||||| || || ||||| ||||| |||
Db         16 ATGGGAAGGGGGAAGATCGAGATCAAGCGGATCGAAAACACCACTAACCGGCAGGTTACT 75

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||||||||| || || ||| | |||||||| |||||||||| || || ||||||
Db         76 TTCTGCAAGCGCCGCAACGGCTTACTCAAGAAGGCCTATGAATTATCTGTCCTATGTGAT 135

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||||| || |||||||||||||||||||||| | ||  | ||  || |||| ||||||| ||
Db        136 GCAGAGGTAGCCCTCATCGTCTTCTCCAGCCGCGGCCGCCTCTACGAATATGCCAACAAC 195

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || || ||   || |||||||||| |||||| | ||| |||   |  | ||    |||
Db        196 AGTGTTAAATCAACAATTGAGAGGTATAAGAAGGCAAGCGCTGATTCCTCCAACACCGGG 255

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | | || ||  ||||| |||||  |  || || ||  |  |||||    || |||
Db        256 TCCGTTTCTGAAGCCAATGCTCAGTTCTACCAGCAAGAATCCTCCAAACTGCATCAGCAG 315

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||   | |||||| ||| | || |||||  || |||||||    ||     | |||||
Db        316 ATTCGTAATTTGCAGAATTCGAACAGGCACATGCTGGGTGAGTCTCTGGGCTCCTTGAAT 375

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              | ||||| ||||||   || ||| || |  | || ||||| || ||  || | |||||
Db        376 TTCAAGGACCTCAAGAGCCTGGAGATTTGGTTAGAGAAAGGCATAAGTAGAATCAGATCC 435

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              |||||||| ||| ||||| ||| |||||||     ||||| | ||||||        |
Db        436 AAAAAGAATGAGCTGTTGTTTGCTGAGATCGAGTATATGCAAAAGAGGGAAATAGACTTG 495

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGC---CACAAC 537
              | || ||  || ||| ||   | |||||||||| | ||| | |  || |   |||
Db        496 CATAATGATAACCAGTATCTCCGAGCAAGGATAGCCGAGAATGAGAGAAACGAACAGCAG 555

Qy        538 ACGAACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCA---------TTCGACTCT 588
              | || | || |  |||     |  ||| |||     ||          |||||||||
Db        556 ATGAGCTTGATGCCAGGAGGAGCTAACTATGAACTCATGCCATCTCAGCAGTTCGACTCT 615

Qy        589 CGAAATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAA 648
              ||||  |  | ||  ||||  | |  |  |   ||   || | |  || || |||
Db        616 AGAAACTATTTCCAACTAAATGGACTGCAACCAAATCAATCATACTCTCGCCAAGACCAA 675

Qy        649 ACAACGCTTCAGCTTGGCTGA 669
              || | |||||  | |  | ||
Db        676 CCAGCTCTTCAATTAGTTTAA 696


RESULT 8
BI311053
LOCUS       BI311053                 754 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  EST5312803 GESD Medicago truncatula cDNA clone pGESD9F6 5' end,
            mRNA sequence.
ACCESSION   BI311053
VERSION     BI311053.1  GI:14985380
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
            ,J. and Fraser,C.M.
  TITLE     ESTs from developing reproductive tissues of Medicago truncatula
  JOURNAL   Unpublished
COMMENT     Contact: Michael A. Grusak
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713-798-7044
            Fax: 713-798-7078
            Email: mgrusak@bcm.tmc.edu
            B397908e
             TIGR sequence name: MTPAP27TK
             More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Medicago truncatula"
```

/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pGESD9F6"
/tissue_type="immature seeds"
/dev_stage="Immature seeds, 11 to 19 days after pollination"
/clone_lib="GESD"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 11 to 19 days after pollination, were harvested from greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propogated in XLOLR cells."

BASE COUNT 274 a 121 c 159 g 200 t
ORIGIN

Query Match 29.8%; Score 271; DB 12; Length 754;
Best Local Similarity 63.0%; Pred. No. 8.9e-64;
Matches 418; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | || ||||||||||||||| |||||||| ||||| || ||  | || || |||
Db         25 ATGGGGAGGGGAAAGATTGAGATAAAAAGGATTGAGAATACAACAAATAGGCAAGTGACT 84

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              || |||||| |  | |||||  |  | ||||| || |||||| |||||||| | ||||||
Db         85 TTTTGCAAGAGAAGGAATGGACTTCTGAAGAAAGCTTATGAACTATCAGTTTTGTGTGAT 144

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || ||||| || ||||| |||||||||||| | || ||||||||||||| | | ||| ||
Db        145 GCTGAAGTTGCTCTCATTGTCTTCTCCAGCCGTGGCAGACTTTATGAATATTCAAACAAC 204

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              | | | | |  || ||||| |||||||| ||  ||||  ||  |
Db        205 AACATAAGGTCAACAATTGATAGGTACAAAAAAGCTTGTTCAGATCATTCAAGTACCACC 264

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | |  | ||   ||  |||||||||  || || ||  |||  || || ||||||||
Db        265 ACTACCACTGAAATTAACGCTCAGTATTATCAGCAAGAATCTGCAAAGCTGAGACAACAA 324

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||  | ||  |||||||| | || |||| |||||||||| |  |  |  | | |
Db        325 ATACAGATGCTGCAAAATTCTAACAGGCACTTGATGGGTGATGCATTAAGCACACTCACT 384

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              | |||||| | ||||| || |||  | |||||||||| ||||||||   || |  ||||
Db        385 GTGAAGGAATTGAAGCAGCTGGAGAATAGACTTGAAAGAGGAATCACTAGAATTAGATCT 444

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||  | |||||| | ||  |  || || ||   | | |||| ||||||        ||
Db        445 AAGAAACATGAGATGCTACTGGCTGAAATTGAATACTTTCAGAAAAGGGAGATTGAGCTG 504

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
                | || |||     |||||| |  ||||||   ||   | |||     |   |   |
Db        505 GAAAATGAAAATCTTTGTCTTCGGACTAAGATAAATGATGTTGAGAGGCTTCCTCAAGTG 564

Qy        541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCTA 600
              |||||| | || | |  ||   ||||| | |  ||||| |  ||||| ||||| |  |
Db        565 AACATGGTTTCTGGACAAGAGCTGAATGCAATCCAAGCATTAGCATCTCGTAATTTTTTC 624

Qy        601 CATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCAG 660
              || ||||  | || || |  |  |  |     |||||  | | |  | | |||
Db        625 AATCCAAACATGATGGAAGATGGTGAAACTTCCTACCATCAATCAGATAAGAAGATTCTT 684

Qy        661 CTT 663
```

```
              | |
Db        685 CAT 687
```

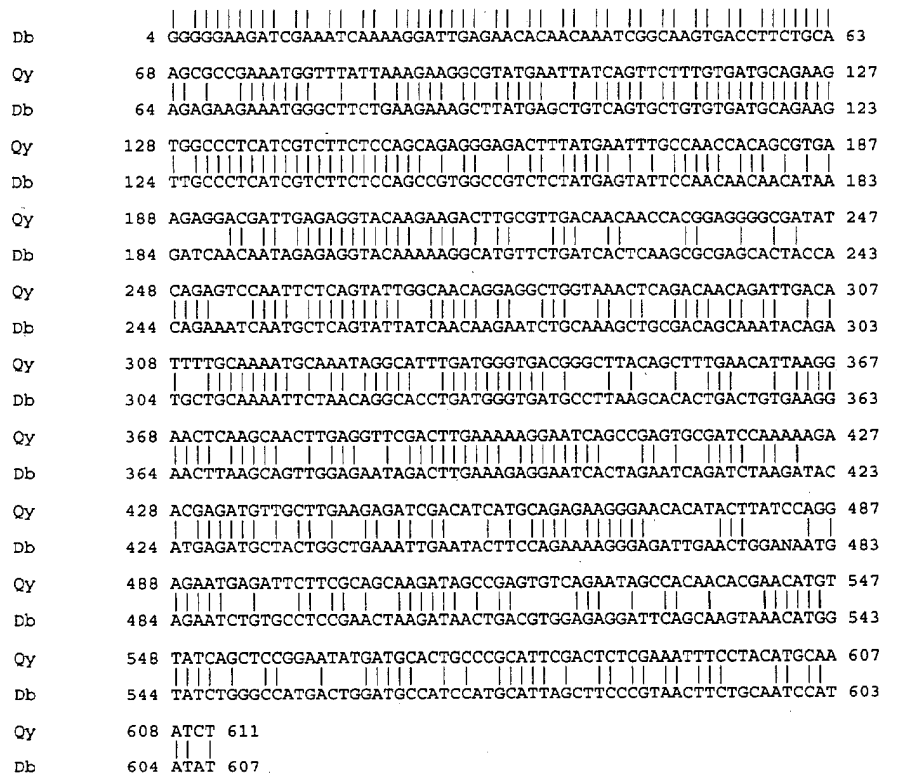
RESULT 9
AW278878
LOCUS AW278878 608 bp mRNA linear EST 02-DEC-2001
DEFINITION sf99h09.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1019-4314 5' similar to TR:Q38836 Q38836 MADS-BOX PROTEIN AGL11. ;, mRNA sequence.
ACCESSION AW278878
VERSION AW278878.1 GI:6667427
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 608)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 799 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 469.

FEATURES Location/Qualifiers
source 1. .608
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-4314"
/tissue_type="immature seed coats of greenhouse grown plants"
/lab_host="DH10B (Gibco BRL)"
/clone_lib="Gm-c1019"
/note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

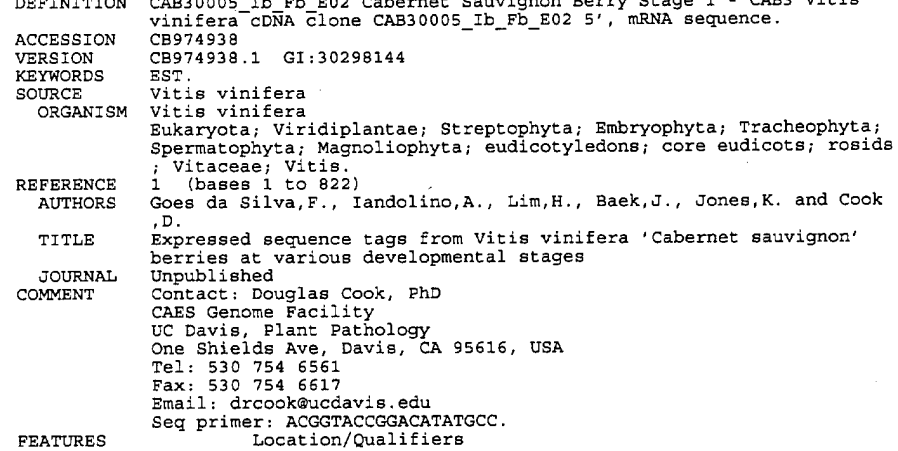
BASE COUNT 209 a 122 c 142 g 134 t 1 others
ORIGIN

Query Match 29.4%; Score 267; DB 9; Length 608;
Best Local Similarity 65.1%; Pred. No. 1e-62;
Matches 393; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
Qy          8 GTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTTTCTGCA 67
```



RESULT 10
CB974938
LOCUS CB974938 822 bp mRNA linear EST 01-MAY-2003
DEFINITION CAB30005_Ib_Fb_E02 Cabernet Sauvignon Berry Stage I - CAB3 Vitis vinifera cDNA clone CAB30005_Ib_Fb_E02 5', mRNA sequence.
ACCESSION CB974938
VERSION CB974938.1 GI:30298144
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 822)
AUTHORS Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook ,D.
TITLE Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon' berries at various developmental stages
JOURNAL Unpublished
COMMENT Contact: Douglas Cook, PhD
CAES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcook@ucdavis.edu
Seq primer: ACGGTACCGGACATATGCC.
FEATURES Location/Qualifiers
source 1. .822
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB30005_Ib_Fb_E02"
/sex="Hermaphrodite"
/dev_stage="Berry stage I"
/lab_host="DH5alpha"
/clone_lib="Cabernet Sauvignon Berry Stage I - CAB3"
/note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2: SfiI; CAB3 is a cDNA library of Vitis vinifera 'Cabernet Sauvignon' Clone 8 berries. Samples were collected after berry set from field-grown vines during stage I of berry growth, 17 days after full bloom. The average berry size was 6 milimeters. Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5'and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT 245 a 183 c 194 g 200 t
ORIGIN



```
Qy        652 ACGCTTCAGCTTGGCTGAACGTTGAAGCGGT 682
              | || || || |  |||     ||  || ||
Db        757 GCTCTCCAACTAGTGTGAGAAGTGGTGCTGT 787


RESULT 11
CB975572
LOCUS       CB975572                824 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  CAB30007_IVb_Fb_A06 Cabernet Sauvignon Berry Stage I - CAB3 Vitis
            vinifera cDNA clone CAB30007_IVb_Fb_A06 5', mRNA sequence.
ACCESSION   CB975572
VERSION     CB975572.1  GI:30298778
KEYWORDS    EST.
SOURCE      Vitis vinifera
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook
            ,D.
  TITLE     Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
            berries at various developmental stages
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: ACGGTACCGGACATATGCC.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Vitis vinifera"
                     /mol_type="mRNA"
                     /cultivar="Cabernet Sauvignon"
                     /db_xref="taxon:29760"
                     /clone="CAB30007_IVb_Fb_A06"
                     /sex="Hermaphrodite"
                     /dev_stage="Berry stage I"
                     /lab_host="DH5alpha"
                     /clone_lib="Cabernet Sauvignon Berry Stage I - CAB3"
                     /note="Organ: Berry; Vector: pDNR; Site_1: SfiI; Site_2:
                     SfiI; CAB3 is a cDNA library of Vitis vinifera 'Cabernet
                     Sauvignon' Clone 8 berries. Samples were collected after
                     berry set from field-grown vines during stage I of berry
                     growth, 17 days after full bloom. The average berry size
                     was 6 milimeters. Sampled vines were located at the
                     University of California, Davis, Experimental Vineyard.
                     cDNAs were made by oligo-dT priming and directionally
                     cloned. 5'and 3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Library was
                     constructed using the Clontech Creator SMART kit and
                     size-selected to contain the 0.5-3 kb size fraction."
BASE COUNT      241 a    195 c    184 g    204 t
ORIGIN

  Query Match           28.9%;  Score 262.6;  DB 14;  Length 824;
  Best Local Similarity   63.4%;  Pred. No. 2e-61;
  Matches  438;  Conservative    0;  Mismatches  244;  Indels    9;  Gaps    2;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              |||||  | |||||||| || || ||| |||| || || || || || ||||| |||||
Db        113 ATGGGAAGAGGGAAGATCGAAATTAAGCGGATCGAGAACACCACTAATCGACAAGTCACC 172

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              ||||| |||||||| || || ||  | || ||||| |||||||||| ||  | ||||||
Db        173 TTCTGTAAGCGCCGCAACGGCTTGCTCAAAAAGGCCTATGAATTATCGGTCTTATGTGAT 232

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              |||||||| || |||||||||||||||||| | ||  | || ||||||| ||| ||| ||

Db        233 GCAGAAGTTGCTCTCATCGTCTTCTCCAGCCGTGGCCGCCTCTATGAATATGCTAACAAC 292

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              || ||||  | |||| || || |||||||| ||     ||   |||   |  | |    ||
Db        293 AGTGTGAGAACGACAATCGAAAGGTACAAAAAAGTGTGTTCTGATTCCTCCAATACCGGA 352

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              |  | || ||  | ||| |||||| ||  || || || ||     ||| |  | |  ||
Db        353 TCTGTTTCTGAAGCTAATGCTCAGTTTTACCAGCAAGAAGCCTCCAAATTGCGCCGTCAA 412

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||      | | || |||   || ||||||  |  |||||||  |  ||||   || |||||
Db        413 ATCAGGGATATACAGAATTTGAACAGGCATATTCTGGGTGAGGCCCTTAGCTCTCTGAAC 472

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | || |||    | | || ||||| |||||| || |  |||||
Db        473 TTTAAGGAACTCAAGAATCTGGAGACCAGGTTGGAGAAAGGTATCAGCAGAATCAGATCC 532

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              |||||||| ||  || || ||| || |||||     ||||| |  |||||        ||
Db        533 AAAAAGAATGAATTGCTGTTTGCTGAAATCGAGTATATGCAAAAGAGGGAGATTGAGCTC 592

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGC---CACAAC 537
              |    ||     |||| ||     ||||||| ||| | |||    |   || |
Db        593 CAAAACAGTAACTTATTTCTGCGAGCACAGATAGCTGAGAATGAGAGAGCCCAACAGCAA 652

Qy        538 ACGAACATGTTATCAGCTCCGGAATATGA------TGCACTGCCCGCATTCGACTCTCGA 591
              | |||| || |  |||   |  | || ||          | | |||  ||| ||||||| |
Db        653 ATGAACCTGATGCCAGGGTCTCAGTACGAGAGCGTGCCGCAGCAGCCATACGACTCTCAA 712

Qy        592 AATTTCCTACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACA 651
              || || || | ||  || ||  | ||  |    || || ||  | |  || || ||||||
Db        713 AACTTGCTCCCTGTTAACCTCCTGGACCCTAATCACCATTACTCTCGCCACGACCAAACA 772

Qy        652 ACGCTTCAGCTTGGCTGAACGTTGAAGCGGT 682
              | || || || |  |||     ||  || ||
Db        773 GCTCTCCAACTAGTGTGAGAAGTGGTGCTGT 803


RESULT 12
CA857928
LOCUS       CA857928                858 bp    mRNA    linear   EST 18-DEC-2002
DEFINITION  EST635183 GLSD Medicago truncatula cDNA clone pGLSD-26B15, mRNA
            sequence.
ACCESSION   CA857928
VERSION     CA857928.1  GI:27232292
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Grusak,M.A., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cheung
            ,F. and Fraser,C.M.
  TITLE     ESTs from late stage developing seeds of Medicago truncatula
  JOURNAL   Unpublished
COMMENT     Contact: Grusak, M.A.
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713 798 7044
            Fax: 713 798 7078
            Email: mgrusak@bcm.tmc.edu
            TIGR sequence name: MTRBM08TK
            More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="A17"
```

/db_xref="taxon:3880"
/clone="pGLSD-26B15"
/tissue_type="Immature seeds"
/dev_stage="25 to 35 days after pollination"
/lab_host="XLOLR"
/clone_lib="GLSD"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 25 to 35 days after pollination, were harvested from greenhouse-grown plants. Seed were removed and separated from pod walls and immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

BASE COUNT 301 a 134 c 179 g 244 t
ORIGIN

Query Match 28.6%; Score 260; DB 14; Length 858;
Best Local Similarity 63.0%; Pred. No. 1.1e-60;
Matches 418; Conservative 0; Mismatches 245; Indels 1; Gaps 1;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGA-GGATTGAAAATACTACGAACCGACAGGTCAC 59
              ||||| | || ||||||||||||||| | ||||||| ||||| || ||  | || || ||
Db         29 ATGGGGAGGGGAAAGATTGAGATAAAAAGGGATTGAGAATACAACAAATAGGCAAGTGAC 88

Qy         60 TTTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGA 119
              ||| |||||| |  | |||||  |  | ||||| || |||||| |||||||| | |||||
Db         89 TTTTTGCAAGAGAAGGAATGGACTTCTGAAGAAAGCTTATGAACTATCAGTTTTGTGTGA 148

Qy        120 TGCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCA 179
              ||| ||||| || ||||| |||||||||||| | || |||||||||||| | | ||| |
Db        149 TGCTGAAGTTGCTCTCATTGTCTTCTCCAGCCGTGGCAGACTTTATGAATATTCAAACAA 208

Qy        180 CAGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGG 239
              || | | | |   || ||||| |||||||| ||  ||||    ||  |
Db        209 CAACATAAGGTCAACAATTGATAGGTACAAAAAAGCTTGTTCAGATCATTCAAGTACCAC 268

Qy        240 GGCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACA 299
                | |  | ||    ||  |||||||||  || || ||  ||| || || ||||||||
Db        269 CACTACCACTGAAATTAACGCTCAGTATTATCAGCAAGAATCTGCAAAGCTGAGACAACA 328

Qy        300 GATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAA 359
               ||  | ||  |||||||| | || ||||| |||||||||||  |  | |  | | |
Db        329 AATACAGATGCTGCAAAATTCTAACAGGCACTTGATGGGTGATGCATTAAGCACACTCAC 388

Qy        360 CATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATC 419
              | |||||| | ||||| || |||  | |||||||||  |||||||    || |  ||||
Db        389 TGTGAAGGAATTGAAGCAGCTGGAGAATAGACTTGAAAGAGGAATCACTAGAATTAGATC 448

Qy        420 CAAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACT 479
               || ||  | |||||| | || |  || || ||   | | |||| ||||||      ||
Db        449 TAAGAAACATGAGATGCTACTGGCTGAAATTGAATACTTTCAGAAAAGGGAGATTGAGCT 508

Qy        480 TATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACAC 539
                | || |||      |||||| |  ||||||   ||   | |||     |   |
Db        509 GGAAAATGAAAATCTTTGTCTTCGGACTAAGATAAATGATGTTGAGAGGCTTCCTCAAGT 568

Qy        540 GAACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATTCGACTCTCGAAATTTCCT 599
              ||||||| | || |  |  ||      ||||| | |  ||||| |  ||||| |||||  |
Db        569 GAACATGGTTTCTGGACAAGAGCTGAATGCAATCCAAGCATTAGCATCTCGTAATTTTTT 628

Qy        600 ACATGCAAATCTAATCGATGCGGCCCATCACTATGCACATCAGGAACAAACAACGCTTCA 659
               || |||| | || || |  |  |     |||||  | | |  | | ||| |
Db        629 CAATCCAAACATGATGGAAGATGGTGAAACTTCCTACCATCAATCAGATAAGAAGATTCT 688

Qy        660 GCTT 663
                | ||
Db        689 TCAT 692
```

RESULT 13
BU045147
LOCUS BU045147 615 bp mRNA linear EST 26-AUG-2002
DEFINITION PP_LEa0021K08f Peach developing fruit mesocarp Prunus persica cDNA clone PP_LEa0021K08f, mRNA sequence.
ACCESSION BU045147
VERSION BU045147.1 GI:22485224
KEYWORDS EST.
SOURCE Prunus persica (peach)
ORGANISM Prunus persica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE 1 (bases 1 to 615)
AUTHORS Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
TITLE Peach Model Genome for Rosaceae
JOURNAL Unpublished
COMMENT Contact: Abbott, A.
Dept of Genetics and Biochemistry
Clemson University
122 Long Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 3060
Fax: 864 656 6879
Email: aalbert@clemson.edu
Total High Quality bases = 585
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 615.
FEATURES Location/Qualifiers
source 1. .615
/organism="Prunus persica"
/mol_type="mRNA"
/cultivar="Loring"
/db_xref="taxon:3760"
/clone="PP_LEa0021K08f"
/tissue_type="Mesocarp"
/lab_host="E. coli"
/clone_lib="Peach developing fruit mesocarp"
/note="Vector: pBluescript II SK(-); Site_1: EcoRI; Site_2: XhoI; authority=Prunus persica L. Batsh; The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis go to http://www.genome.clemson.edu/projects/peach. To order this clone go to http://www.genome.clemson.edu/orders"
BASE COUNT 219 a 123 c 144 g 129 t
ORIGIN

Query Match 28.4%; Score 258.6; DB 13; Length 615;
Best Local Similarity 67.6%; Pred. No. 2.2e-60;
Matches 363; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||| | || ||||| ||||| ||| |||| ||||| || || ||||| || ||||| |
Db         65 TGGGTAGGGGAAAGATCGAGATCAAGCGGATCGAAAACACAACAAACCGTCAAGTCACCT 124

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              ||||||| | || ||||| ||  | |||||||| |||||| | || || || |||||||
Db        125 TCTGCAAAAGGCGCAATGGGTTGCTCAAGAAGGCCTATGAACTCTCTGTACTCTGTGATG 184

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              |||| || || ||||| || ||||| | | | ||  |||||||||| | |||||||  |||
Db        185 CAGAGGTTGCTCTCATAGTTTTCTCTAACCGTGGCCGACTTTATGAGTATGCCAACAACA 244

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||     || |||||||||||||||| | ||||  ||   |  |    ||
Db        245 GTGTTAAAGAAACAATTGAGAGGTACAAGAAGGCATGCGCAGAGTCTACAAATACCGGAT 304

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  | | | |||||||  |  ||||| || ||||  |||||  |   ||||
Db        305 CTGTTTCCGAAGCTAGTACTCAGTACTACCAACAAGAAGCTGCGAAACTGCGCGCTCAGA 364
```

```
Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              | |  | ||||||  ||  |||  ||||||  |||||||  ||       |  ||  |||| |
Db        365 TAGGGAATTTGCAGAACTCAAGCAGGCATATGATGGGCGAGTCATTGAGTTCTATGAATA 424

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | || || || ||| | || |||  |  ||| || ||||||||||  | ||  |  ||||||
Db        425 TGAAAGACCTGAAGAATCTGGAGAGTAAACTAGAGAAAGGAATCAACAGAATCAGATCCA 484

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              | |||||  ||  | |||  |||  ||  ||  ||   |||||||  |||||||        |
Db        485 AGAAGAATGAACTCTTGTTTGCCGAAATTGAGTACATGCAGAAAAGGGAAATTGACTTGC 544

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACA 538
                  |  | ||  ||  | ||  ||    ||||||||  |||  | |||  |||||  | |
Db        545 ATAACAACAACCAGCTCCTACGAGCAAAGATAGCTGAGAATGAGAGGAGCCAGCAAA 601


RESULT 14
BU043610
LOCUS       BU043610                631 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  PP_LEa0016J23f Peach developing fruit mesocarp Prunus persica cDNA
            clone PP_LEa0016J23f, mRNA sequence.
ACCESSION   BU043610
VERSION     BU043610.1  GI:22483687
KEYWORDS    EST.
SOURCE      Prunus persica (peach)
  ORGANISM  Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
  TITLE     Peach Model Genome for Rosaceae
  JOURNAL   Unpublished
COMMENT     Contact: Abbott, A.
            Dept of Genetics and Biochemistry
            Clemson University
            122 Long Hall, Clemson University, Clemson, SC 29634, USA
            Tel: 864 656 3060
            Fax: 864 656 6879
            Email: aalbert@clemson.edu
            Total High Quality bases = 586
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 631.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Prunus persica"
                     /mol_type="mRNA"
                     /cultivar="Loring"
                     /db_xref="taxon:3760"
                     /clone="PP_LEa0016J23f"
                     /tissue_type="Mesocarp"
                     /lab_host="E. coli"
                     /clone_lib="Peach developing fruit mesocarp"
                     /note="Vector: pBluescript II SK(-); Site_1: EcoRI;
                     Site_2: XhoI; authority=Prunus persica L. Batsh; The
                     sequence has been trimmed to remove vector sequence and
                     contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis go to
                     http://www.genome.clemson.edu/projects/peach. To order
                     this clone go to http://www.genome.clemson.edu/orders"
BASE COUNT      222 a    124 c    153 g    132 t
ORIGIN

  Query Match           28.4%;  Score 258.6;  DB 13;  Length 631;
  Best Local Similarity   67.6%;  Pred. No. 2.2e-60;
  Matches  363;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              |||| | || |||| |||| ||| |||| |||| || || |||| || |||| |
Db         64 TGGGTAGGGGAAAGATCGAGATCAAGCGGATCGAAAACACAACAAACCGTCAAGTCACCT 123

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || |||| ||  | |||||||| |||||  | || || || |||||||
Db        124 TCTGCAAAAGGCGCAATGGGTTGCTCAAGAAGGCCTATGAACTCTCTGTACTCTGTGATG 183

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              |||| || || |||| || |||| | | | ||  |||||||||| | ||||||| |||
Db        184 CAGAGGTTGCTCTCATAGTTTTCTCTAACCGTGGCCGACTTTATGAGTATGCCAACAACA 243

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||     || |||||||||||||||||| | ||||  ||      |    ||
Db        244 GTGTTAAAGAAACAATTGAGAGGTACAAGAAGGCATGCGCAGAGTCTACAAATACCGGAT 303

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              |  | || ||  | | | |||||||  |  ||||| || ||||  |||||  |  ||||
Db        304 CTGTTTCCGAAGCTAGTACTCAGTACTACCAACAAGAAGCTGCGAAACTGCGCGCTCAGA 363

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              | |  | ||||||  ||  |||  ||||||  |||||||  ||       |  ||  |||| |
Db        364 TAGGGAATTTGCAGAACTCAAGCAGGCATATGATGGGCGAGTCATTGAGTTCTATGAATA 423

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | || || || ||| | || |||  |  ||| || ||||||||||  | ||  |  ||||||
Db        424 TGAAAGACCTGAAGAATCTGGAGAGTAAACTAGAGAAAGGAATCAACAGAATCAGATCCA 483

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              | |||||  ||  | |||  ||  ||  ||   |||||||  |||||||        |
Db        484 AGAAGAATGAACTCTTGTTTGCCGAAATTGAGTACATGCAGAAAAGGGAAATTGACTTGC 543

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACA 538
              | |  | ||  ||  | ||  ||    ||||||||  |||  | |||  |||||  | |
Db        544 ATAACAACAACCAGCTCCTACGAGCAAAGATAGCTGAGAATGAGAGGAGCCAGCAAA 600


RESULT 15
AW184799
LOCUS       AW184799                649 bp    mRNA    linear   EST 02-DEC-2001
DEFINITION  se82f12.y1 Gm-c1023 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1023-1224 5' similar to TR:Q38836 Q38836 MADS-BOX PROTEIN
            AGL11. ;, mRNA sequence.
ACCESSION   AW184799
VERSION     AW184799.1  GI:6454186
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
```

/clone="GENOME SYSTEMS CLONE ID: Gm-c1023-1224"
/tissue_type="seed coats of greenhouse grown plants"
/lab_host="DH10B"
/clone_lib="Gm-c1023"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This cDNA library was constructed from mRNA isolated from seed coats (100-200mgs) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complimentary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restrictions site. Sal I linkers adapers were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 231 a 123 c 156 g 139 t
ORIGIN

Query Match 28.4%; Score 257.8; DB 9; Length 649;
Best Local Similarity 65.2%; Pred. No. 3.8e-60;
Matches 379; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

```
Qy    1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db   68 ATGGGAAGGGGGAAGATCGAAATCAAAAGGATTGAGAACACAACAAATCGGCAAGTGACC 127

Qy   61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db  128 TTCTGCAAGAGAAGAAATGGGCTTCTGAAGAAAGCTTATGAGCTGTCAGTGCTGTGTGAT 187

Qy  121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db  188 GCAGAAGTTGCCCTCATCGTCTTCTCCAGCCGTGGCCGTCTCTATGAGTATTCCAACAAC 247

Qy  181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db  248 AACATAAGATCAACAATAGAGAGGTACAAAAAGGCATGTTCTGATCACTCAAGCGCGAGC 307

Qy  241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db  308 ACTACCACAGAAATCAATGCTCAGTATTATCAGCAAGAATCTGCAAAGCTGCGACAGCAA 367

Qy  301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
Db  368 ATACAGATGCTGCAAAATTCTAACAGGCACCTGATGGGTGATGCCTTAAGCACACTGACT 427

Qy  361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
Db  428 GTGAAGGAACTTAAGCAGTTGGAGAATAGACTTGAAAGAGGAATCACTAGAATCAGATCT 487

Qy  421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
Db  488 AAGAAACATGAGATGCTACTGGCTGAAATTGAATACTTCCAGAAAAGGGAGATTGAACTG 547

Qy  481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGAATAGCCACAACACG 540
Db  548 GAAAATGAAAATCTTTGCCTCCGAACTAAGATAACTGACGTGGAGAGGATTCAGCAAGTA 607

Qy  541 AACATGTTATCAGCTCCGGAATATGATGCACTGCCCGCATT 581
Db  608 AACATGGTTTCTGGGCCAGAACTGAATGCCATTCAAGCATT 648
```

Search completed: January 22, 2004, 00:54:24
Job time : 1821.1 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:11:45 ; Search time 306.264 Seconds
(without alignments)
10461.606 Million cell updates/sec

Title: US-09-936-869-3
Perfect score: 909
Sequence: 1 atgggtcgtgggaagattga.....cgcatgttatcttttctgttg 909

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

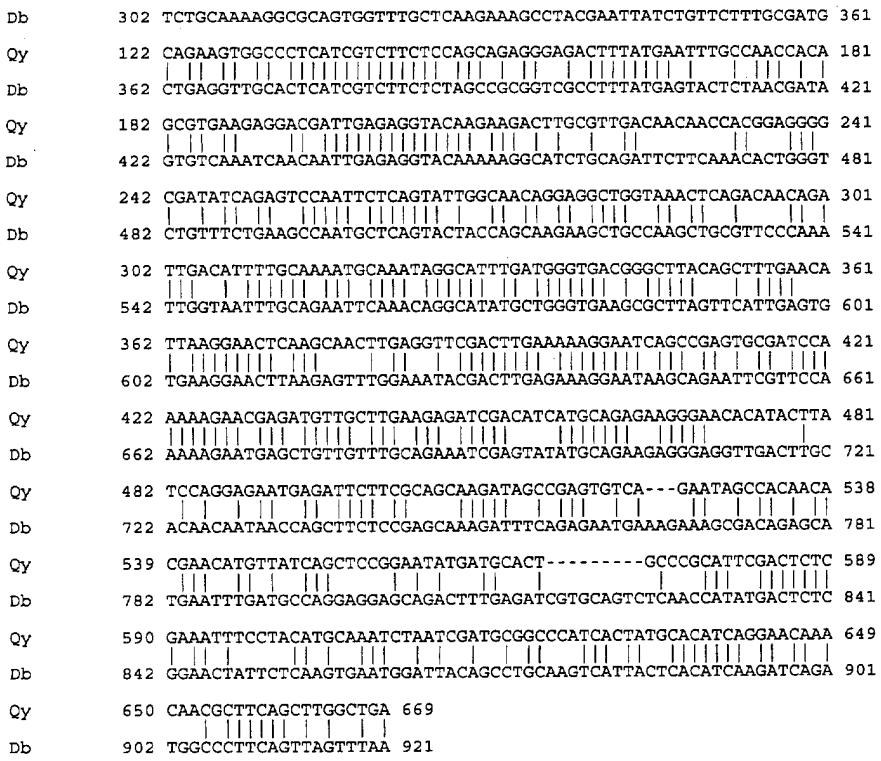
Result No.	Score	% Query Match	Length	DB	ID	Description
1	260	28.6	1219	15	US-10-104-580-10	Sequence 10, Appl
2	259.8	28.6	723	15	US-10-104-580-11	Sequence 11, Appl
3	255.6	28.1	959	10	US-09-978-730-5	Sequence 5, Appli
4	255.6	28.1	959	10	US-09-978-729A-5	Sequence 5, Appli
5	255.6	28.1	959	10	US-09-981-087A-5	Sequence 5, Appli
6	255.6	28.1	959	10	US-09-978-382A-5	Sequence 5, Appli
7	255.6	28.1	959	11	US-09-978-740A-5	Sequence 5, Appli
8	248.2	27.3	714	15	US-10-104-580-15	Sequence 15, Appl
9	248.2	27.3	1159	15	US-10-104-580-14	Sequence 14, Appl
10	245	27.0	896	10	US-09-978-730-3	Sequence 3, Appli
11	245	27.0	896	10	US-09-978-729A-3	Sequence 3, Appli
12	245	27.0	896	10	US-09-981-087A-3	Sequence 3, Appli
13	245	27.0	896	10	US-09-978-382A-3	Sequence 3, Appli
14	245	27.0	896	11	US-09-978-740A-3	Sequence 3, Appli
15	230.2	25.3	905	15	US-10-286-264-160	Sequence 160, App
16	201.8	22.2	795	13	US-10-259-165-775	Sequence 775, App
17	171.2	18.8	1280	10	US-09-970-624-1	Sequence 1, Appli
18	164.6	18.1	1038	11	US-09-819-142-3	Sequence 3, Appli
19	164.6	18.1	1038	11	US-09-934-455-31	Sequence 31, Appl
c 20	160.4	17.6	845	9	US-09-770-445-644	Sequence 644, App
21	157.8	17.4	1210	11	US-09-819-142-17	Sequence 17, Appl
22	157.8	17.4	1210	11	US-09-934-455-403	Sequence 403, App
23	157.8	17.4	1210	15	US-10-295-403-69	Sequence 69, Appl
24	154.2	17.0	824	13	US-10-259-165-161	Sequence 161, App
25	152.8	16.8	645	10	US-09-938-842A-1776	Sequence 1776, Ap
26	152.8	16.8	645	10	US-09-853-450-33	Sequence 33, Appl
27	148.6	16.3	959	11	US-09-819-142-13	Sequence 13, Appl
28	148.6	16.3	959	11	US-09-934-455-355	Sequence 355, App
29	145.4	16.0	738	13	US-10-259-165-415	Sequence 415, App
30	145.4	16.0	741	13	US-10-259-165-45	Sequence 45, Appl
31	141.6	15.6	1062	10	US-09-978-730-1	Sequence 1, Appli
32	141.6	15.6	1062	10	US-09-978-729A-1	Sequence 1, Appli
33	141.6	15.6	1062	10	US-09-981-087A-1	Sequence 1, Appli
34	141.6	15.6	1062	10	US-09-978-382A-1	Sequence 1, Appli
35	141.6	15.6	1062	11	US-09-978-740A-1	Sequence 1, Appli
36	138.8	15.3	485	13	US-10-259-165-538	Sequence 538, App
37	138	15.2	1345	10	US-09-853-450-7	Sequence 7, Appli
38	137.6	15.1	735	13	US-10-259-165-713	Sequence 713, App
39	136.8	15.0	666	10	US-09-938-842A-2442	Sequence 2442, Ap
40	136.4	15.0	499	13	US-10-259-165-539	Sequence 539, App
41	134.8	14.8	1098	11	US-09-819-142-15	Sequence 15, Appl
42	134.8	14.8	1098	11	US-09-934-455-357	Sequence 357, App
43	133.8	14.7	747	10	US-09-853-450-27	Sequence 27, Appl
44	133.6	14.7	906	16	US-10-278-536-195	Sequence 195, App
45	133.2	14.7	705	10	US-09-938-842A-2404	Sequence 2404, Ap

ALIGNMENTS

RESULT 1
US-10-104-580-10
; Sequence 10, Application US/10104580
; Publication No. US20030033628A1
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 62486
; CURRENT APPLICATION NUMBER: US/10/104,580
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 09/410,464
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/287,700
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: 60/080,851
; PRIOR FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1219
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(921)
US-10-104-580-10

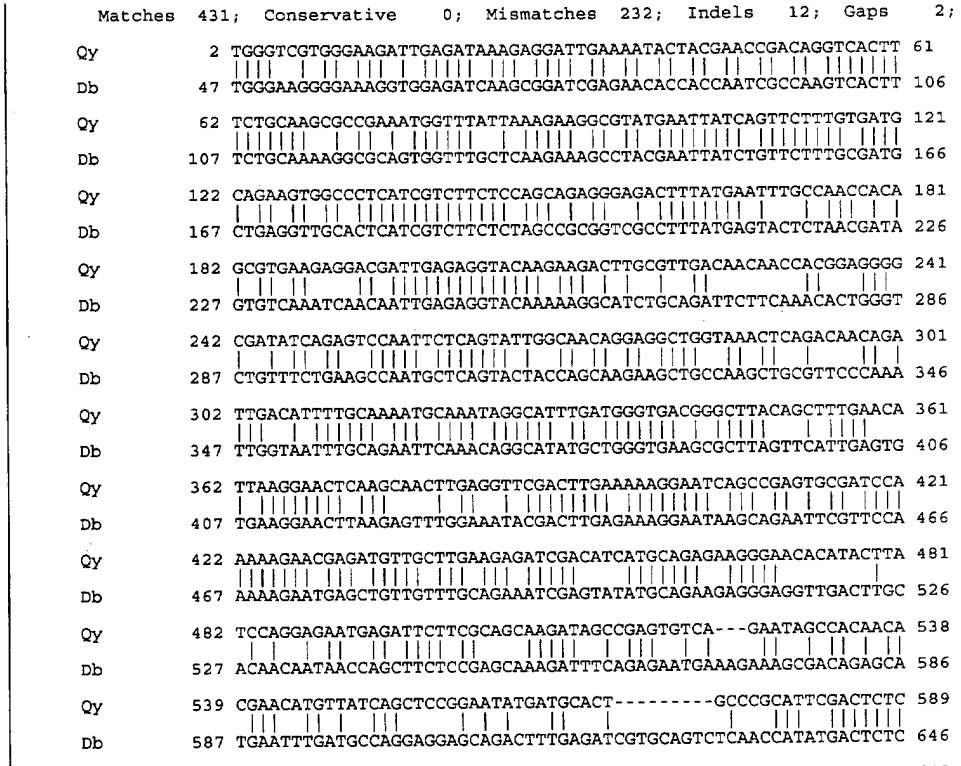
Query Match 28.6%; Score 260; DB 15; Length 1219;
Best Local Similarity 63.7%; Pred. No. 1.4e-65;
Matches 433; Conservative 0; Mismatches 235; Indels 12; Gaps 2;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db        242 TGGGAAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 301

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || | ||||||  | ||||| || || |||||||| |||||||| ||||
```



RESULT 2
US-10-104-580-11
; Sequence 11, Application US/10104580
; Publication No. US20030033628A1
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 62486
; CURRENT APPLICATION NUMBER: US/10/104,580
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 09/410,464
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/287,700
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: 60/080,851
; PRIOR FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(723)
US-10-104-580-11

Query Match 28.6%; Score 259.8; DB 15; Length 723;
Best Local Similarity 63.9%; Pred. No. 1.1e-65;



RESULT 3
US-09-978-730-5
; Sequence 5, Application US/09978730
; Patent No. US20020129403A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000920US
; CURRENT APPLICATION NUMBER: US/09/978,730
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 959
; TYPE: DNA

; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(818)
; OTHER INFORMATION: AGAMOUS-LIKE 5 (AGL5)
US-09-978-730-5

Query Match 28.1%; Score 255.6; DB 10; Length 959;
Best Local Similarity 68.3%; Pred. No. 2.3e-64;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||  ||   | ||||||| |||||||||||||| ||  || |||||||| ||  || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || |||||||| | ||||| || |||||  | || ||  | |||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||  ||  ||  ||  ||   || |||||||||    ||||   |  ||  ||  ||  | |||||  ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||  ||||    | ||  ||  || ||||||||||   |||||    |||  |    ||      |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||   | |||  |||||||  |   ||  ||  |||||    |||||||  |    ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||        | ||  |||    ||  ||  ||   |  | |||||     |||     | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              ||||||||||||||  |  ||||||    |  |||||  ||||||||||  ||  ||  ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              ||  |||  ||||||||||   |||  ||||||  ||    ||||||  |  |||||||    |    ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                |  ||  ||    |  |||  |||   ||||||   |  ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640
```

RESULT 4
US-09-978-729A-5
; Sequence 5, Application US/09978729A
; Patent No. US20020178465A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000950US
; CURRENT APPLICATION NUMBER: US/09/978,729A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(818)
; OTHER INFORMATION: AGAMOUS-LIKE 5 (AGL5)
US-09-978-729A-5

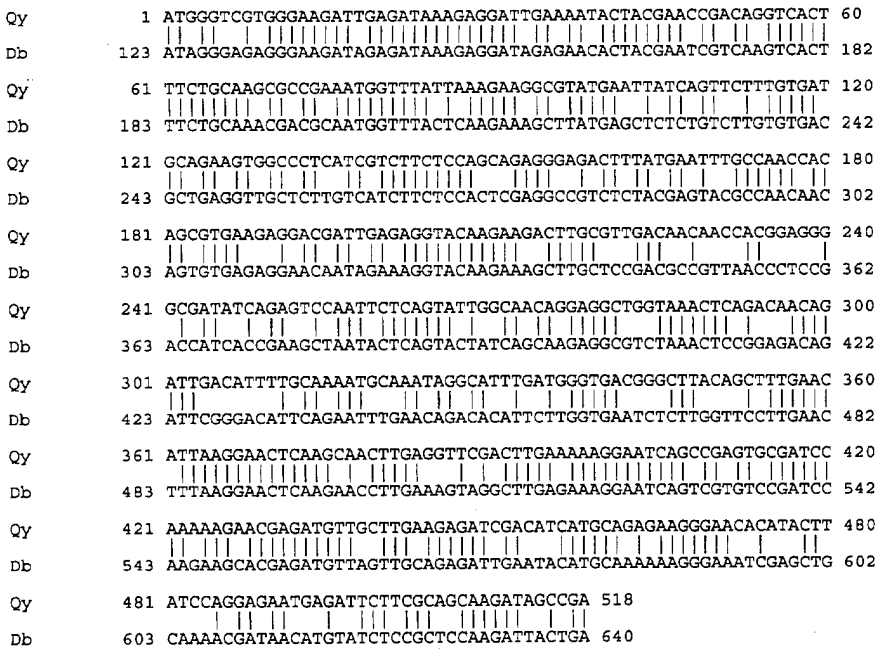
Query Match 28.1%; Score 255.6; DB 10; Length 959;
Best Local Similarity 68.3%; Pred. No. 2.3e-64;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||  ||   | ||||||| |||||||||||||| ||  || |||||||| ||  || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || |||||||| | ||||| || |||||  | || ||  | |||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              ||  ||  ||  ||  ||   || |||||||||    ||||   |  ||  ||  ||  | |||||  ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||  ||||    | ||  ||  || ||||||||||   |||||    |||  |    ||      |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||   | |||  |||||||  |   ||  ||  |||||    |||||||  |    ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||        | ||  |||    ||  ||  ||   |  | |||||     |||     | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              ||||||||||||||  |  ||||||    |  |||||  ||||||||||  ||  ||  ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              ||  |||  ||||||||||   |||  ||||||  ||    ||||||  |  |||||||    |    ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                |  ||  ||    |  |||  |||   ||||||   |  ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640
```

RESULT 5
US-09-981-087A-5
; Sequence 5, Application US/09981087A
; Patent No. US20020178466A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: Farrandiz, Cristina
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000940US
; CURRENT APPLICATION NUMBER: US/09/981,087A
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(818)
; OTHER INFORMATION: AGAMOUS-LIKE 5 (AGL5)

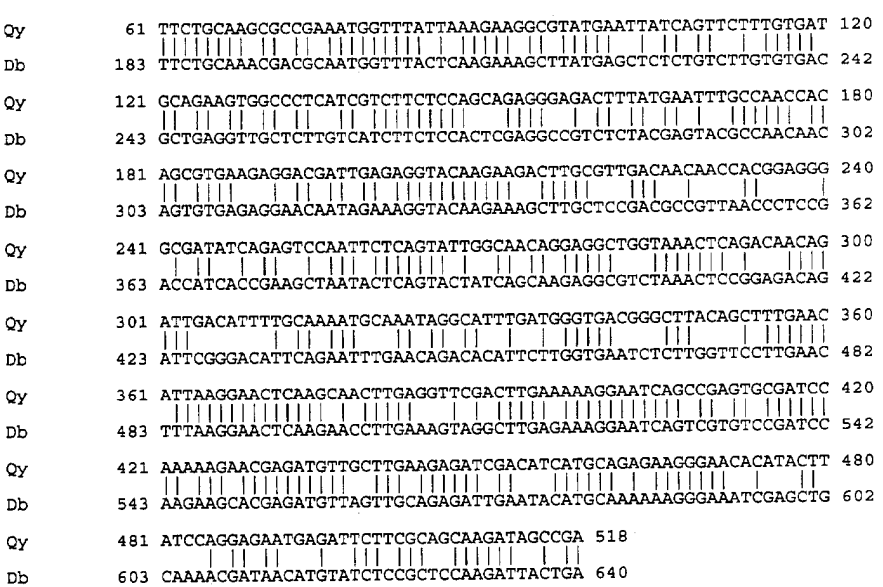
US-09-981-087A-5

Query Match 28.1%; Score 255.6; DB 10; Length 959;
Best Local Similarity 68.3%; Pred. No. 2.3e-64;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;



RESULT 6
US-09-978-382A-5
; Sequence 5, Application US/09978382A
; Publication No. US20020194647A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000930US
; CURRENT APPLICATION NUMBER: US/09/978,382A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(818)
; OTHER INFORMATION: AGAMOUS-LIKE 5 (AGL5)
US-09-978-382A-5

Query Match 28.1%; Score 255.6; DB 10; Length 959;
Best Local Similarity 68.3%; Pred. No. 2.3e-64;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;



RESULT 7
US-09-978-740A-5
; Sequence 5, Application US/09978740A
; Publication No. US20030005481A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000960US
; CURRENT APPLICATION NUMBER: US/09/978,740A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(818)
; OTHER INFORMATION: AGAMOUS-LIKE 5 (AGL5)
US-09-978-740A-5

Query Match 28.1%; Score 255.6; DB 11; Length 959;
Best Local Similarity 68.3%; Pred. No. 2.3e-64;
Matches 354; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
              ||  ||   | ||||||| ||||||||||||||| || || |||||||| || || ||||||
Db        123 ATAGGGAGAGGGAAGATAGAGATAAAGAGGATAGAGAACACTACGAATCGTCAAGTCACT 182

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
              |||||||| || || |||||||| | ||||| || |||||   | || ||  | ||||||
Db        183 TTCTGCAAACGACGCAATGGTTTACTCAAGAAAGCTTATGAGCTCTCTGTCTTGTGTGAC 242

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
              || || || ||  || |||||||||    |||| | || || || |  |||||| ||
Db        243 GCTGAGGTTGCTCTTGTCATCTTCTCCACTCGAGGCCGTCTCTACGAGTACGCCAACAAC 302

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
              ||  ||||   | || || || ||||||||||  ||||   |||  |   ||      |
Db        303 AGTGTGAGAGGAACAATAGAAAGGTACAAGAAAGCTTGCTCCGACGCCGTTAACCCTCCG 362

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
              | ||  | ||  | ||| |||||||  |  || || |||||   |||||| |   ||||
Db        363 ACCATCACCGAAGCTAATACTCAGTACTATCAGCAAGAGGCGTCTAAACTCCGGAGACAG 422

Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              |||      | || |||   || || ||  |  | |||||    |||    | ||||||
Db        423 ATTCGGGACATTCAGAATTTGAACAGACACATTCTTGGTGAATCTCTTGGTTCCTTGAAC 482

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              |||||||||||||| | |||||   | | ||||| |||||||||| || || ||||||
Db        483 TTTAAGGAACTCAAGAACCTTGAAAGTAGGCTTGAGAAAGGAATCAGTCGTGTCCGATCC 542

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| |||||||||  ||| |||||| ||   |||||| | |||||||   |   ||
Db        543 AAGAAGCACGAGATGTTAGTTGCAGAGATTGAATACATGCAAAAAAGGGAAATCGAGCTG 602

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
                | || ||   |  ||| |||  ||||||  | ||
Db        603 CAAAACGATAACATGTATCTCCGCTCCAAGATTACTGA 640
```

RESULT 8
US-10-104-580-15
; Sequence 15, Application US/10104580
; Publication No. US20030033628A1
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 62486
; CURRENT APPLICATION NUMBER: US/10/104,580
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 09/410,464
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/287,700
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: 60/080,851
; PRIOR FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(714)
US-10-104-580-15

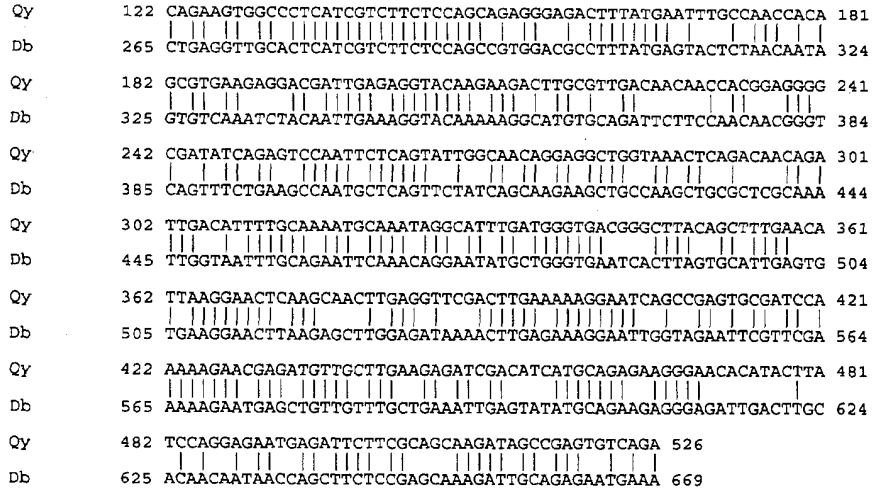
Query Match 27.3%; Score 248.2; DB 15; Length 714;
Best Local Similarity 67.0%; Pred. No. 2.9e-62;
Matches 352; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db         47 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 106

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || ||||||||  | ||||| || |||||||||| |||||||| ||||
Db        107 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 166

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
              | || || || ||||||||||||||||| | ||  | |||||||| |   | ||| | |
Db        167 CTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAATA 226

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
              | || ||  || ||||| |||||||| ||| | || |  ||      | ||   |||
Db        227 GTGTCAAATCTACAATTGAAAGGTACAAAAAGGCATGTGCAGATTCTTCCAACAACGGGT 286

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
              | | || ||  ||||| |||||  |  || || || |||| || ||  |   || |
Db        287 CAGTTTCTGAAGCCAATGCTCAGTTCTATCAGCAAGAAGCTGCCAAGCTGCGCTCGCAAA 346

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
              |||  | |||||| ||| |||| ||| || || |||||||    |||| || ||||
Db        347 TTGGTAATTTGCAGAATTCAAACAGGAATATGCTGGGTGAATCACTTAGTGCATTGAGTG 406

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
              | |||||||| |||    | ||| |   |||||| |||||||  |  || | || || |
Db        407 TGAAGGAACTTAAGAGCTTGGAGATAAAACTTGAGAAAGGAATTGGTAGAATTCGTTCGA 466

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
              ||||||| ||| ||||| |||  || || ||    |||||||  |||||          |
Db        467 AAAAGAATGAGCTGTTGTTTGCTGAAATTGAGTATATGCAGAAGAGGGAGATTGACTTGC 526

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGAGTGTCAGA 526
                | |  | ||  || |||| ||    ||||| || |||  | | |
Db        527 ACAACAATAACCAGCTTCTCCGAGCAAAGATTGCAGAGAATGAAA 571
```

RESULT 9
US-10-104-580-14
; Sequence 14, Application US/10104580
; Publication No. US20030033628A1
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; TITLE OF INVENTION: poplar and other plant species.
; FILE REFERENCE: 62486
; CURRENT APPLICATION NUMBER: US/10/104,580
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: 09/410,464
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/287,700
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: 60/080,851
; PRIOR FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (99)..(815)
US-10-104-580-14

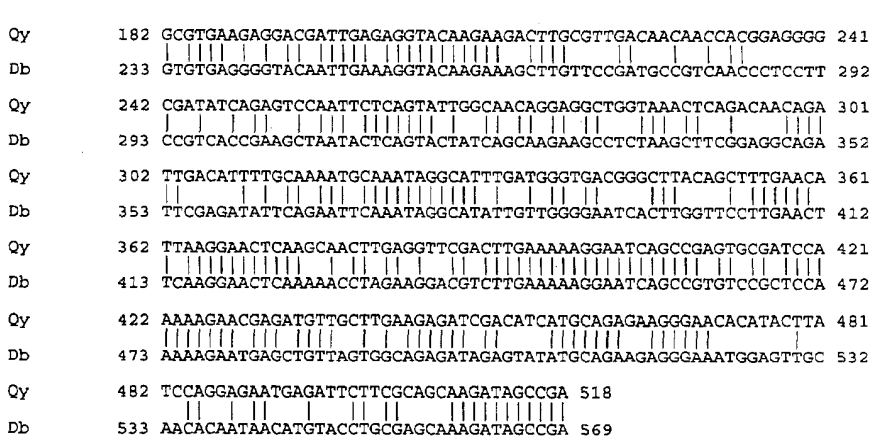
Query Match 27.3%; Score 248.2; DB 15; Length 1159;
Best Local Similarity 67.0%; Pred. No. 3.9e-62;
Matches 352; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
              ||||  | || ||| | ||||| ||| |||| || || || || || || || |||||||
Db        145 TGGGGAGGGGAAAGGTGGAGATCAAGCGGATCGAGAACACCACCAATCGCCAAGTCACTT 204

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
              |||||||  | || ||||||||  | ||||| || |||||||||| |||||||| ||||
Db        205 TCTGCAAAAGGCGGAATGGTTTGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGATG 264
```



RESULT 10
US-09-978-730-3
; Sequence 3, Application US/09978730
; Patent No. US20020129403A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000920US
; CURRENT APPLICATION NUMBER: US/09/978,730
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 896
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(753)
; OTHER INFORMATION: AGAMOUS-LIKE 1 (AGL1)
US-09-978-730-3

Query Match 27.0%; Score 245; DB 10; Length 896;
Best Local Similarity 67.1%; Pred. No. 2.9e-61;
Matches 347; Conservative 0; Mismatches 170; Indels 0; Gaps 0;



RESULT 11
US-09-978-729A-3
; Sequence 3, Application US/09978729A
; Patent No. US20020178465A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljegren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000950US
; CURRENT APPLICATION NUMBER: US/09/978,729A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 896
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(753)
; OTHER INFORMATION: AGAMOUS-LIKE 1 (AGL1)
US-09-978-729A-3

Query Match 27.0%; Score 245; DB 10; Length 896;
Best Local Similarity 67.1%; Pred. No. 2.9e-61;
Matches 347; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```
Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301

Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361

Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421

Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481

Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518

Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569
```

```
RESULT 12
US-09-981-087A-3
; Sequence 3, Application US/09981087A
; Patent No. US20020178466A1
; GENERAL INFORMATION:
;  APPLICANT: Yanofsky, Martin F.
;  APPLICANT:  Liljegren, Sarah
;  APPLICANT:  Farrandiz, Cristina
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 19452A-000940US
;  CURRENT APPLICATION NUMBER: US/09/981,087A
;  CURRENT FILING DATE:  2002-05-21
;  PRIOR APPLICATION NUMBER: US 60/090,649
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: US 09/339,998
;  PRIOR FILING DATE: 1999-06-25
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 896
;   TYPE: DNA
;   ORGANISM: Arabidopsis sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(753)
;   OTHER INFORMATION: AGAMOUS-LIKE 1 (AGL1)
US-09-981-087A-3

  Query Match           27.0%;  Score 245;  DB 10;  Length 896;
  Best Local Similarity   67.1%;  Pred. No. 2.9e-61;
  Matches  347;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61

Db         53 TAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACAAATCGTCAAGTTACTT 112

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121

Db        113 TCTGCAAACGACGCAATGGTCTTCTCAAGAAAGCTTATGAACTCTCTGTCTTGTGTGATG 172

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181

Db        173 CCGAAGTTGCCCTCGTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAACA 232

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241

Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301

Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361

Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421

Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481

Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518

Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569
```

```
RESULT 13
US-09-978-382A-3
; Sequence 3, Application US/09978382A
; Publication No. US20020194647A1
; GENERAL INFORMATION:
;  APPLICANT: Yanofsky, Martin F.
;  APPLICANT:  Liljegren, Sarah
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 19452A-000930US
;  CURRENT APPLICATION NUMBER: US/09/978,382A
;  CURRENT FILING DATE:  2002-06-18
;  PRIOR APPLICATION NUMBER: US 60/090,649
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: US 09/339,998
;  PRIOR FILING DATE: 1999-06-25
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 896
;   TYPE: DNA
;   ORGANISM: Arabidopsis sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(753)
;   OTHER INFORMATION: AGAMOUS-LIKE 1 (AGL1)
US-09-978-382A-3

  Query Match           27.0%;  Score 245;  DB 10;  Length 896;
  Best Local Similarity   67.1%;  Pred. No. 2.9e-61;
  Matches  347;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61

Db         53 TAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACAAATCGTCAAGTTACTT 112

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121

Db        113 TCTGCAAACGACGCAATGGTCTTCTCAAGAAAGCTTATGAACTCTCTGTCTTGTGTGATG 172

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181

Db        173 CCGAAGTTGCCCTCGTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAACA 232

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241

Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301

Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352
```

```
Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569

RESULT 14
US-09-978-740A-3
; Sequence 3, Application US/09978740A
; Publication No. US20030005481A1
; GENERAL INFORMATION:
;  APPLICANT: Yanofsky, Martin F.
;  APPLICANT:  Liljegren, Sarah
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 19452A-000960US
;  CURRENT APPLICATION NUMBER: US/09/978,740A
;  CURRENT FILING DATE:  2001-10-15
;  PRIOR APPLICATION NUMBER: US 60/090,649
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: US 09/339,998
;  PRIOR FILING DATE: 1999-06-25
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 896
;   TYPE: DNA
;   ORGANISM: Arabidopsis sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(753)
;   OTHER INFORMATION: AGAMOUS-LIKE 1 (AGL1)
US-09-978-740A-3

  Query Match             27.0%;  Score 245;  DB 11;  Length 896;
  Best Local Similarity   67.1%;  Pred. No. 2.9e-61;
  Matches  347;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          2 TGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACTT 61
Db         53 TAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACAAATCGTCAAGTTACTT 112

Qy         62 TCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGATG 121
Db        113 TCTGCAAACGACGCAATGGTCTTCTCAAGAAAGCTTATGAACTCTCTGTCTTGTGTGATG 172

Qy        122 CAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCACA 181
Db        173 CCGAAGTTGCCCTCGTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAACA 232

Qy        182 GCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGGG 241
Db        233 GTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGATGCCGTCAACCCTCCTT 292

Qy        242 CGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAGA 301
Db        293 CCGTCACCGAAGCTAATACTCAGTACTATCAGCAAGAAGCCTCTAAGCTTCGGAGGCAGA 352

Qy        302 TTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAACA 361
Db        353 TTCGAGATATTCAGAATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAACT 412

Qy        362 TTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCCA 421
Db        413 TCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATCAGCCGTGTCCGCTCCA 472

Qy        422 AAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTTA 481
Db        473 AAAAGAATGAGCTGTTAGTGGCAGAGATAGAGTATATGCAGAAGAGGGAAATGGAGTTGC 532

Qy        482 TCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCGA 518
Db        533 AACACAATAACATGTACCTGCGAGCAAAGATAGCCGA 569

RESULT 15
US-10-286-264-160
; Sequence 160, Application US/10286264
; Publication No. US20030093837A1
; GENERAL INFORMATION:
;  APPLICANT: Keddie, James
;  APPLICANT:  Riechmann, Jose-Luis
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Zhang, James
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Adam, Luc
;  APPLICANT:  Broun, Pierre
;  APPLICANT:  Reuber, Lynne
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Samaha, Raymond
;  TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION
;  FILE REFERENCE: MBI-008
;  CURRENT APPLICATION NUMBER: US/10/286,264
;  CURRENT FILING DATE:  2002-11-01
;  PRIOR APPLICATION NUMBER: 60/125,814
;  PRIOR FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 165
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 160
;   LENGTH: 905
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G146
US-10-286-264-160

  Query Match             25.3%;  Score 230.2;  DB 15;  Length 905;
  Best Local Similarity   64.8%;  Pred. No. 6.6e-57;
  Matches  357;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy          1 ATGGGTCGTGGGAAGATTGAGATAAAGAGGATTGAAAATACTACGAACCGACAGGTCACT 60
Db         20 ATGGGAAGAGGAAAGATAGAAATAAAGAGGATAGAGAACTCAACAAATCGACAAGTGACG 79

Qy         61 TTCTGCAAGCGCCGAAATGGTTTATTAAAGAAGGCGTATGAATTATCAGTTCTTTGTGAT 120
Db         80 TTTTGCAAAAGAAGAAATGGACTTCTGAAGAAAGCCTATGAGCTTTCGGTCCTTTGCGAT 139

Qy        121 GCAGAAGTGGCCCTCATCGTCTTCTCCAGCAGAGGGAGACTTTATGAATTTGCCAACCAC 180
Db        140 GCAGAAGTTGCGCTCATTGTTTTCTCCACTCGTGGCCGTCTCTATGAATACGCCAATAAC 199

Qy        181 AGCGTGAAGAGGACGATTGAGAGGTACAAGAAGACTTGCGTTGACAACAACCACGGAGGG 240
Db        200 AACATAAGATCAACCATTGAGAGGTACAAGAAAGCTTGTTCTGATAGCACCAACACTAGC 259

Qy        241 GCGATATCAGAGTCCAATTCTCAGTATTGGCAACAGGAGGCTGGTAAACTCAGACAACAG 300
Db        260 ACTGTCCAAGAAATCAATGCCGCGTACTATCAACAAGAATCTGCTAAGCTGAGACAACAG 319
```

```
Qy        301 ATTGACATTTTGCAAAATGCAAATAGGCATTTGATGGGTGACGGGCTTACAGCTTTGAAC 360
              ||  | |   | |||||  | || ||| || |||||||| |||    | |   | || |
Db        320 ATCCAAACGATTCAAAACTCCAACAGGAATCTGATGGGAGACTCTTTGAGTTCCTTAAGT 379

Qy        361 ATTAAGGAACTCAAGCAACTTGAGGTTCGACTTGAAAAAGGAATCAGCCGAGTGCGATCC 420
              | |||||||| || ||| |||||  ||| ||||| |||| ||| |  | | | |||
Db        380 GTCAAGGAACTAAAACAAGTTGAGAATCGCCTTGAGAAAGCTATCTCTAGGATCAGGTCC 439

Qy        421 AAAAAGAACGAGATGTTGCTTGAAGAGATCGACATCATGCAGAGAAGGGAACACATACTT 480
              || ||| | ||| || | || |  || ||||| | |  ||||| |||||        |||
Db        440 AAGAAGCATGAGTTGCTTCTAGTTGAAATCGAAAACGCGCAGAAAAGGGAGATTGAGCTT 499

Qy        481 ATCCAGGAGAATGAGATTCTTCGCAGCAAGATAGCCG-AGTGTCAGAATAGCCACAACAC 539
              | | |||||       |||  | |  ||| |||| | ||||       || | |||||||
Db        500 GACAATGAGAACATCTATCTAAGAACTAAGGTAGCAGAAGTGGAGAGGTATCAACAACAC 559

Qy        540 GAACATGTTAT 550
              | |||   ||
Db        560 CATCATCAAAT 570
```

Search completed: January 22, 2004, 01:20:39
Job time : 310.264 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:02:36 ; Search time 2018.55 Seconds
(without alignments)
12185.050 Million cell updates/sec

Title: US-09-936-869-5
Perfect score: 1012
Sequence: 1 atcgaattaaagtcaatggc.....ttctatggtaaaaaaaaaaa 1012

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	634	62.6	635	9	AV825172	AV825172 AV825172
c 2	521	51.5	530	9	AI995637	AI995637 701676626
3	520	51.4	520	9	AV527174	AV527174 AV527174
4	491	48.5	607	9	AV832175	AV832175 AV832175
c 5	490	48.4	490	9	AB038725	AB038725 AB038725
c 6	457.6	45.2	527	9	AV520789	AV520789 AV520789
c 7	420.4	41.5	422	9	AV788022	AV788022 AV788022
c 8	417.4	41.2	419	9	AV820692	AV820692 AV820692
c 9	416.4	41.1	418	9	AV819666	AV819666 AV819666
c 10	402	39.7	402	9	AV819464	AV819464 AV819464
c 11	391.4	38.7	394	14	Z26559	Z26559 ATTS1651 Ve
12	378.8	37.4	570	13	BQ791108	BQ791108 E3954 Chi
c 13	375.6	37.1	400	9	AV816618	AV816618 AV816618
14	371	36.7	371	9	AV527256	AV527256 AV527256
15	365.2	36.1	453	14	R90379	R90379 16734 Lambd
c 16	359.4	35.5	361	9	AV532040	AV532040 AV532040
17	349.8	34.6	849	12	BM359116	BM359116 GA__Ea001
18	343	33.9	351	10	BE524429	BE524429 M49H2STM
19	339.4	33.5	684	9	AI729386	AI729386 BNLGHi132
20	331.2	32.7	771	14	CB292632	CB292632 UCRCS01_0
21	328	32.4	685	10	BG585152	BG585152 EST486915
c 22	319.6	31.6	865	14	CB292631	CB292631 UCRCS01_0
23	315.4	31.2	941	9	AI729649	AI729649 BNLGHi138
24	311.6	30.8	630	13	BQ405390	BQ405390 GA__Ed008
25	306.4	30.3	690	9	AW039494	AW039494 EST281775
26	306.2	30.3	664	14	CA826107	CA826107 R72F05 tw
27	305	30.1	627	13	BU825514	BU825514 UK109TF10
28	304	30.0	804	14	CA801284	CA801284 sau03e09.
29	303.2	30.0	392	14	T23025	T23025 5033 Lambda
30	302	29.8	741	14	CA799030	CA799030 sat75c05.
31	301.4	29.8	632	10	BE191140	BE191140 sn85f11.y
32	301.4	29.8	646	9	AW775880	AW775880 EST334945
33	294.4	29.1	645	10	BG597684	BG597684 EST496362
34	290.2	28.7	658	10	BG455225	BG455225 NF103D11P
35	285.4	28.2	351	14	T44541	T44541 7804 Lambda
36	285.4	28.2	594	12	BI420700	BI420700 LjNEST60e
37	285.4	28.2	674	10	BG581199	BG581199 EST482931
38	280	27.7	646	10	BE124916	BE124916 EST393951
39	279.2	27.6	651	13	BQ119665	BQ119665 EST605241
40	279	27.6	567	10	BF113095	BF113095 EST440685
41	278.4	27.5	590	10	BE435686	BE435686 EST406764
42	278.4	27.5	613	10	BE020073	BE020073 sm38f09.y
43	274.6	27.1	670	10	AW980478	AW980478 EST391631
44	274.4	27.1	702	9	AW223027	AW223027 EST299838
45	271.4	26.8	654	9	AW775566	AW775566 EST334631

ALIGNMENTS

RESULT 1
AV825172
LOCUS AV825172 635 bp mRNA linear EST 01-APR-2002
DEFINITION AV825172 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-88-B05 5', mRNA sequence.
ACCESSION AV825172
VERSION AV825172.1 GI:19867232
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 635)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998).This clone is in a modified pBluescript vector as a SstI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL06-88-B05"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL6"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      154 a    138 c    145 g    198 t
ORIGIN
```

```
  Query Match             62.6%;  Score 634;  DB 9;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-135;
  Matches  634;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 AATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGG 74
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGG 61

Qy         75 AGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCA 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCA 121

Qy        135 ATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAG 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAG 181

Qy        195 AGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGG 254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 AGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGG 241

Qy        255 TTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGAT 314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGAT 301

Qy        315 GGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGG 374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGG 361

Qy        375 TGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGT 434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGT 421

Qy        435 TTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGT 494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGT 481

Qy        495 CACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGG 554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGG 541

Qy        555 AGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAG 614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 AGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAG 601

Qy        615 AGATGCAATCGATGAAATACGTATATGCTTCTAT 648
              |||||||||||||||||||||||||||||||||
Db        602 AGATGCAATCGATGAAATACGTATATGCTTCTAT 635
```

```
RESULT 2
AI995637/c
LOCUS       AI995637                530 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  701676626 A. thaliana, Columbia Col-0, inflorescence-1 Arabidopsis
            thaliana cDNA clone 701676626, mRNA sequence.
ACCESSION   AI995637
VERSION     AI995637.1  GI:5842542
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Columbia Col-0"
                     /db_xref="taxon:3702"
                     /clone="701676626"
                     /tissue_type="inflorescence"
                     /dev_stage="4 - 7 weeks"
                     /clone_lib="A. thaliana, Columbia Col-0, inflorescence-1"
                     /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA
                     library was derived from untreated inflorescence tissue
                     from Arabidopsis  thaliana, Columbia Col-0, at 4 - 7
                     weeks. Plants were grown in 1:1:1 peat
                     moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C
                     under constant light,  and watered with fertilizer. cDNA
                     synthesis was initiated using a NotI-oligo(dT)  primer.
                     Double-stranded cDNA was blunted, ligated to SalI adaptors
                     , digested with  NotI, size-selected, and cloned into the
                     NotI and SalI sites of the pSPORT vector."
BASE COUNT      163 a    108 c    100 g    158 t      1 others
ORIGIN
```

```
  Query Match             51.5%;  Score 521;  DB 9;  Length 530;
  Best Local Similarity   99.8%;  Pred. No. 1.6e-109;
  Matches  521;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        404 GGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTA 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 GGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTA 463

Qy        464 CCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATG 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 CCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATG 403

Qy        524 TTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCAT 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 TTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCAT 343

Qy        584 TTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCT 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 TTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCT 283

Qy        644 TCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGT 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 TCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGT 223
```

```
Qy        704 CATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTC 763
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTC 163

Qy        764 TGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCT 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 TGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCT 103

Qy        824 TCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAG 883
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 TCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAG 43

Qy        884 ATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTG 925
              |||||||||||||||||||||||||||||||||||| |||||
Db         42 ATTGGCTTATTTCGTTCTATTGGATATGTATCATCANTACTG 1
```

RESULT 3
AV527174
LOCUS AV527174 520 bp mRNA linear EST 01-SEP-2000
DEFINITION AV527174 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone APZ30g03R 5', mRNA sequence.
ACCESSION AV527174
VERSION AV527174.1 GI:8686702
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 520)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .520
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APZ30g03R"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 118 a 114 c 121 g 167 t
ORIGIN

Query Match 51.4%; Score 520; DB 9; Length 520;
Best Local Similarity 100.0%; Pred. No. 2.6e-109;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          6 ATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGC 65
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGC 60

Qy         66 ATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCT 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCT 120

Qy        126 ATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGC 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGC 180

Qy        186 TTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTA 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTA 240

Qy        246 TAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTC 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTC 300

Qy        306 AACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATC 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATC 360

Qy        366 ATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTC 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTC 420

Qy        426 TTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAA 485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCTCCTGTTTTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAA 480

Qy        486 GCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTT 525
              |||||||||||||||||||||||||||||||||||||||
Db        481 GCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTT 520
```

RESULT 4
AV832175
LOCUS AV832175 607 bp mRNA linear EST 01-APR-2002
DEFINITION AV832175 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-05-B04 5', mRNA sequence.
ACCESSION AV832175
VERSION AV832175.1 GI:19874235
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 607)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998).cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.
FEATURES Location/Qualifiers
source 1. .607
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL11-05-B04"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL11"
/note="Site_1: BamHI; Site_2: SalI; subjected to various treatments (dehydration, cold, high salt, ABA, heat and UV

```
                ). Dark-grown plants"
BASE COUNT      148 a    134 c    134 g    190 t      1 others
ORIGIN

  Query Match           48.5%;  Score 491;  DB 9;  Length 607;
  Best Local Similarity   93.5%;  Pred. No. 1.3e-102;
  Matches  548;  Conservative    0;  Mismatches    1;  Indels   37;  Gaps    2;

Qy          1 ATCGAATTAAAGTCAATGGC-GTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGC 59
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         22 ATCGAATTAAAGTCAATGGCAGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGC 81

Qy         60 CGGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTT 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 CGGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTT 141

Qy        120 CGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAA 179
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 CGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAA 201

Qy        180 AAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCC 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 AAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCC 261

Qy        240 TGACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGA 299
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 TGACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGA 321

Qy        300 GATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCC 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 GATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCC 381

Qy        360 ATCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGG---------- 406
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 ATCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGTTCCATGCTTA 441

Qy        407 ----------------------AGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGA 443
                                     |||||||||||||||||||||||||||||||||||
Db        442 ATCTCACAACCTTTTGCATAACAAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGA 501

Qy        444 TGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGT 503
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 TGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGT 561

Qy        504 CCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCT 549
              |||||||||||||||||||||||||||||||||||| |||||
Db        562 CCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGNATCCT 607


RESULT 5
AB038725/c
LOCUS       AB038725                490 bp    mRNA    linear   EST 14-NOV-2000
DEFINITION  AB038725 Arabidopsis thaliana Above-ground organ from two to
            six-week old plants Columbia Arabidopsis thaliana cDNA clone
            APZ30g03_f 3', mRNA sequence.
ACCESSION   AB038725
VERSION     AB038725.1  GI:7212552
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZ30g03_f"
                     /tissue_type="aboveground organ"
                     /dev_stage="two to six-week old plants"
                     /clone_lib="Arabidopsis thaliana Above-ground organ from
                     two to six-week old plants Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      147 a     99 c     87 g    157 t
ORIGIN

  Query Match           48.4%;  Score 490;  DB 9;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-102;
  Matches  490;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        485 AGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGT 544
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 AGCATAATGTCACGGATGTCCTTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGT 431

Qy        545 ATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGG 604
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 ATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGG 371

Qy        605 TTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCA 664
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 TTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCA 311

Qy        665 GGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTT 724
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 GGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTT 251

Qy        725 TGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAG 784
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 TGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAG 191

Qy        785 AAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACAT 844
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 AAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACAT 131

Qy        845 GTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATT 904
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 GTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATT 71

Qy        905 GGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAA 964
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 GGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAA 11

Qy        965 AATCCATAAG 974
              ||||||||||
Db         10 AATCCATAAG 1


RESULT 6
AV520789/c
LOCUS       AV520789                527 bp    mRNA    linear   EST 01-SEP-2000
DEFINITION  AV520789 Arabidopsis thaliana aboveground organs two to six-week
            old Arabidopsis thaliana cDNA clone APZ32g06F 3', mRNA sequence.
ACCESSION   AV520789
VERSION     AV520789.1  GI:8680316
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZ32g06F"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /clone_lib="Arabidopsis thaliana aboveground organs two to
                     six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      157 a    111 c    101 g    158 t
ORIGIN

  Query Match           45.2%;  Score 457.6;  DB 9;  Length 527;
  Best Local Similarity   94.1%;  Pred. No. 6.4e-95;
  Matches  497;  Conservative    0;  Mismatches   29;  Indels    2;  Gaps    2;

Qy        392 GGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGTATA 451
              ||||||||||||||||||  |||||||||| |||| ||| ||||||||||||||||||||
Db        527 GGGGCCACGAGTGGGAGAACAATGGGACTTG-TCTTTTCCGGTTTTTCATGATGAGTATA 469

Qy        452 ATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCCTTTATC 511
              ||||||||| || | || ||||||| |||||||||||||||||||||| ||||||||||
Db        468 ATTACTTCCCTAGCCCAGTTAATCTGTACTTGAAGCATAATGTCACGAATGTCCTTTATC 409

Qy        512 AAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCA 571
              || |||| |||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db        408 AAACTGGATATGTTGCTTCCAACAGTGAAAAGTATCCTATAGTAGGTATCGTAACAGCCA 349

Qy        572 TTCAGAATGCATTTCATATCACCCCTGAAGTGGT-TTGCAAAAGAGATGCAATCGATGAA 630
              ||||| |||||||||||||||||||||||||||  |||||||||||||||||||||||||
Db        348 TTCAGTATGCATTTCATATCACCCCTGAAGTGGTCATGCAAAAGAGATGCAATCGATGAA 289

Qy        631 ATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTG 690
              ||||||||||||||| || |||||||| ||||||||||||||||||||||||||||||||
Db        288 ATACGTATATGCTTCAATTAAGATTTTGAGCCCAGGGACTGTGTTGGTTCACAAGATTTG 229

Qy        691 ACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGT 750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 ACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGT 169

Qy        751 GAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGG 810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 GAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGG 109

Qy        811 AGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACC 870
              |||||||||||||| ||||||||||||||||||||| |||   |||   ||||||||||
Db        108 AGCTTTGTTATCTTATGAGAGACAATACATACATGTTTCTCCCGTTCCCACTTTACTACC 49

Qy        871 AAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATC 918
              |||||||||||||||||| |||||||||| ||||||||||||||||
Db         48 AAAACCTATAAAGATTGGCATATTTCGTTCGATTGGATATGTATCATC 1


RESULT 7
AV788022/c
LOCUS       AV788022                422 bp    mRNA    linear   EST 28-MAR-2002
DEFINITION  AV788022 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-77-N15 3',
            mRNA sequence.
ACCESSION   AV788022
VERSION     AV788022.1  GI:19806812
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL06-77-N15"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL6"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      127 a     83 c     72 g    140 t
ORIGIN

  Query Match           41.5%;  Score 420.4;  DB 9;  Length 422;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-86;
  Matches  421;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        574 CAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATA 633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATA 363

Qy        634 CGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTGACA 693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGATTTGACA 303

Qy        694 TCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGTGAA 753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGATGGTGAA 243

Qy        754 GCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGC 813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGC 183

Qy        814 TTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAA 873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAA 123

Qy        874 ACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCA 933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCA 63
```

```
Qy        934 AGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTGAGT 993
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTGAGT 3

Qy        994 TC 995
              |
Db          2 CC 1
```

RESULT 8
AV820692/c

```
LOCUS       AV820692                419 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV820692 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-11-M08 3',
            mRNA sequence.
ACCESSION   AV820692
VERSION     AV820692.1  GI:19862671
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL11-11-M08"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL11"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to various
                     treatments (dehydration, cold, high salt, ABA, heat and UV
                     ). Dark-grown plants"
BASE COUNT      128 a     85 c     72 g    134 t
ORIGIN
```

```
  Query Match             41.2%;  Score 417.4;  DB 9;  Length 419;
  Best Local Similarity   99.8%;  Pred. No. 1.2e-85;
  Matches  418;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        542 AGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAG 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 AGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAG 360

Qy        602 TGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGC 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGC 300

Qy        662 CCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAA 721
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        299 CCAGGGACTGTGTTGGTTCACAAGATTTGAAATCTAGAAAGTCATGCCCCAAGTACGTAA 240

Qy        722 GTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAA 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 GTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAA 180

Qy        782 GAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATA 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATA 120

Qy        842 CATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCT 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 CATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCT 60

Qy        902 ATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         59 ATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 1
```

RESULT 9
AV819666/c

```
LOCUS       AV819666                418 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV819666 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-06-E06 3',
            mRNA sequence.
ACCESSION   AV819666
VERSION     AV819666.1  GI:19861609
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL11-06-E06"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL11"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to various
                     treatments (dehydration, cold, high salt, ABA, heat and UV
                     ). Dark-grown plants"
BASE COUNT      128 a     85 c     74 g    131 t
ORIGIN
```

```
  Query Match             41.1%;  Score 416.4;  DB 9;  Length 418;
  Best Local Similarity   99.8%;  Pred. No. 2e-85;
  Matches  417;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy        543 GTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGT 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 GTATCCTCTAGGAGGTATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGT 359

Qy        603 GGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCC 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 GGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCC 299

Qy        663 CAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAG 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 CAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAG 239

Qy        723 TTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAG 782
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 TTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAG 179

Qy        783 AGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATAC 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 AGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATAC 119

Qy        843 ATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTA 902
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        118 ATGTCTCTGATGTTGTAACTTTACTACCCAAACCTATAAAGATTGGCTTATTTCGTTCTA 59

Qy        903 TTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 TTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 1
```

RESULT 10
AV819464/c

```
LOCUS       AV819464                402 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV819464 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-05-B04 3',
            mRNA sequence.
ACCESSION   AV819464
VERSION     AV819464.1  GI:19861405
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL11-05-B04"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL11"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to various
                     treatments (dehydration, cold, high salt, ABA, heat and UV
                     ). Dark-grown plants"
BASE COUNT      123 a     80 c     70 g    129 t
ORIGIN
```

```
  Query Match             39.7%;  Score 402;  DB 9;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-82;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        559 ATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGAT 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 ATCGTAACAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGAT 343

Qy        619 GCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGT 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 GCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGT 283

Qy        679 TCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 TCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACG 223

Qy        739 CCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATC 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATC 163

Qy        799 GGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGT 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 GGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGT 103

Qy        859 AACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATC 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 AACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATC 43

Qy        919 ATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 960
              ||||||||||||||||||||||||||||||||||||||||||
Db         42 ATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATC 1
```

RESULT 11
Z26559/c

```
LOCUS       Z26559                  394 bp    mRNA    linear   EST 31-MAY-1995
DEFINITION  ATTS1651 Versailles-VC Arabidopsis thaliana cDNA clone VCVDH08 3'
            similar to Ribonuclease (RNS2), mRNA sequence.
ACCESSION   Z26559
VERSION     Z26559.1  GI:404227
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   CNRS.
  TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
  JOURNAL   Unpublished
COMMENT     Contact: Desprez T., Amselem J., Chiapello H., Rouze P., Caboche
            M., Hofte H.
            INRA Versailles
            Laboratoire de Biologie Cellulaire
            Route de Saint-Cyr,78026 Versailles Cedex,France
            Email: thierry@versailles.inra.fr.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="ecotype Columbia"
                     /db_xref="taxon:3702"
                     /clone="VCVDH08"
                     /tissue_type="whole seedlings"
                     /dev_stage="in vitro-grown etiolated seedlings,5 days old"
                     /clone_lib="Versailles-VC"
                     /note="Vector: pBluescript"
```

```
BASE COUNT      117 a     80 c     63 g    133 t      1 others
ORIGIN

  Query Match           38.7%;  Score 391.4;  DB 14;  Length 394;
  Best Local Similarity   99.5%;  Pred. No. 1.1e-79;
  Matches  392;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        599 AAGTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTTCTATAAAGATTTTA 658
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 AAGNGGTTTGCAAAAGAGATGCAATCGATGAAATACGTATATGCTCCTATAAAGATTTTA 335

Qy        659 AGCCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACG 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 AGCCCAGGGACTGTGTTGGTTCACAAGATTTGACATCTAGAAAGTCATGCCCCAAGTACG 275

Qy        719 TAAGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAG 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 TAAGTTTGCCGGAATACACGCCATTAGATGGTGAAGCTATGGTTCTGAAGATGCCAACAG 215

Qy        779 AAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATAC 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 AAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATAC 155

Qy        839 ATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGT 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 ATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGT 95

Qy        899 TCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGA 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 TCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGA 35

Qy        959 TCAGAAAATCCATAAGAAGATATCAACATTTGAG 992
              |||||||||||||||||||||||||||||||||
Db         34 TCAGAAAATCCATAAGAAGATATCAACATTTGAG 1


RESULT 12
BQ791108
LOCUS       BQ791108                 570 bp    mRNA    linear   EST 30-JUL-2002
DEFINITION  E3954 Chinese cabbage etiolated seedling library Brassica rapa
            subsp. pekinensis cDNA clone E3954, mRNA sequence.
ACCESSION   BQ791108
VERSION     BQ791108.1  GI:22006086
KEYWORDS    EST.
SOURCE      Brassica rapa subsp. pekinensis
  ORGANISM  Brassica rapa subsp. pekinensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Ryu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.
  TITLE     Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
            (2002)
  JOURNAL   Unpublished
COMMENT     Contact: Lim, C.O.
            Plant Molecular Biology & Biotechnology Research Centre
            Gyeongsang National University
            #900 Gazwa-dong, Jinju 660-701, Korea
            Tel: 82 55 751 6255
            Fax: 82 55 759 9363
            Email: colim@nongae.gsnu.ac.kr
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Brassica rapa subsp. pekinensis"
                     /mol_type="mRNA"
                     /cultivar="Jangwon"
                     /db_xref="taxon:51351"
                     /clone="E3954"
                     /tissue_type="Etiolated seedling"
                     /lab_host="XL-1 Blue"
                     /clone_lib="Chinese cabbage etiolated seedling library"
                     /note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

BASE COUNT      122 a     149 c     133 g     166 t
ORIGIN

  Query Match           37.4%;  Score 378.8;  DB 13;  Length 570;
  Best Local Similarity   89.6%;  Pred. No. 9.5e-77;
  Matches  407;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         76 GACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAA 135
              |||||| ||| |||||||| || |||||||||| |||||||||||| |||| |||||||
Db        117 GACGTCGTCGTACTCAATCGCTCGCAGAGGGAGTTTGATTATTTCGCTCTCTCTCTCCAA 176

Qy        136 TGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGA 195
              ||||| |||||||| |||||||||||||||||| |||||||||||||||||| ||| |||
Db        177 TGGCCCGGAACCTACTGCCGTGGAACTCGCCATTGCTGCTCCAAAAACGCTTGCTGTAGA 236

Qy        196 GGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGT 255
              ||||| ||  |||||||||||||| || |||||  |||||| |||||||||| ||| |||
Db        237 GGCTCGGACACTCCAACTCAATTCACCATCCATGGACTATGGCCGGACTATAACGACGGT 296

Qy        256 TCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATG 315
              || |||||||||||||| |||||||| ||||| || |||||||||| || |||||||||
Db        297 TCCTGGCCTTCATGTTGCTATCGATCCGACTTCAATGAGAAGGAGATCTCTACGTTGATG 356

Qy        316 GATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGT 375
              |||||  | || |||||||||| |||||||| |||||||||| ||||| ||| |||||
Db        357 GATGGCATGGATAAGTACTGGCCCAGTCTCAGCTGTGGTTCTCCGTCATCTTGCCATGGT 416

Qy        376 GGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTT 435
              || ||||| |||||||||||| |||||||||||||||||||||||||||| |||||||||
Db        417 GGAAAAGGATCATTTTGGGGCCATGAGTGGGAGAAACATGGGACTTGCTCTTCTCCTGTT 476

Qy        436 TTTCATGATGAGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTC 495
              ||| |||||||||| ||||||| ||||||| |||||||| ||||| ||||||||||||
Db        477 ATTCGTGATGAGTATAGTTACTTCATTACCACGCTTAATCTATACTTAAAGCATAATGTC 536

Qy        496 ACGGATGTCCTTTATCAAGCTGGCTATGTTGCTT 529
              |||||||||||||||||||||||||||||||||
Db        537 ACGGATGTCCTTTATCAAGCTGGCTATGTTGCTT 570


RESULT 13
AV816618/c
LOCUS       AV816618                 400 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV816618 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-91-F09 3',
            mRNA sequence.
ACCESSION   AV816618
VERSION     AV816618.1  GI:19858474
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
```

details.
FEATURES Location/Qualifiers
source 1. .400
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-91-F09"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL9"
/note="Site_1: BamHI; Site_2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
BASE COUNT 127 a 77 c 64 g 132 t
ORIGIN

Query Match 37.1%; Score 375.6; DB 9; Length 400;
Best Local Similarity 99.0%; Pred. No. 5e-76;
Matches 378; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy        628 GAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGAT 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 GAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCACAAGAT 341

Qy        688 TTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGAT 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TTGACATCTAGAAAGTCATGCCCCAAGTACGTAAGTTTGCCGGAATACACGCCATTAGAT 281

Qy        748 GGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGAT 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 GGTGAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGAT 221

Qy        808 GGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACT 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 GGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACT 161

Qy        868 ACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGT 927
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 ACCCAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGT 101

Qy        928 AAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACAT 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 AAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACAT 41

Qy        988 TTGAGTTCTATGGTAAAAAAAA 1009
              ||||||||||||||||  ||
Db         40 TTGAGTTCTATGGTAAATTGAA 19
```

RESULT 14
AV527256
LOCUS AV527256 371 bp mRNA linear EST 01-SEP-2000
DEFINITION AV527256 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone APZ32g06R 5', mRNA sequence.
ACCESSION AV527256
VERSION AV527256.1 GI:8686784
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 371)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .371
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="APZ32g06R"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 82 a 85 c 90 g 114 t
ORIGIN

Query Match 36.7%; Score 371; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 5.7e-75;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         10 AAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTT 69
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTT 60

Qy         70 GCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCT 129
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCT 120

Qy        130 CTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGC 189
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAAAACGCTTGC 180

Qy        190 TGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAAC 249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAAC 240

Qy        250 GATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACG 309
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACG 300

Qy        310 TTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGC 369
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGC 360

Qy        370 AATGGTGGGAA 380
              ||||||||||||
Db        361 AATGGTGGGAA 371
```

RESULT 15
R90379
LOCUS R90379 453 bp mRNA linear EST 30-DEC-1997
DEFINITION 16734 Lambda-PRL2 Arabidopsis thaliana cDNA clone 185O18T7, mRNA sequence.
ACCESSION R90379
VERSION R90379.1 GI:957919
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 453)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel ,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729

```
   PUBMED   7846151
COMMENT     Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="185018T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA.  "
BASE COUNT       97 a     89 c    115 g    134 t     18 others
ORIGIN

  Query Match             36.1%;  Score 365.2;  DB 14;  Length 453;
  Best Local Similarity   89.2%;  Pred. No. 1.3e-73;
  Matches  395;  Conservative    0;  Mismatches   45;  Indels    3;  Gaps    1;

Qy          1 ATCGAATTAAAGTCAATGGCGTCACGTTTATGTCTTCTCCTTCTCGTTGCGTGTATCGCC 60
              ||||| ||||||||||||||||||||||||||| || ||||||||||||||||||||
Db          1 ATCGATTTAAAGTCAATGGCGTCACGTTTATGTNTTNTCCTTCTCGTTGCGTGTATCGCC 60

Qy         61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCATTTGCCGGAGACGTCATCGAACTCAATCGATCTCAGAGGGAGTTCGATTATTTC 120

Qy        121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTGTTGCTCCAAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        121 GCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGGAACTCGCCATTNTTGCTCCAAA 180

Qy        181 AACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACGCTTGTTGCAGAGGCTCCGATGCTCCAACTCAATTCACAATTCATGGGTTATGGCCT 240

Qy        241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACTATAACGATGGTTCGTGGCCTTCATGTTGTTATCGATCTGACTTTAAAGAGAAGGAG 300

Qy        301 ATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCA 360
              ||||||||| |||||||||| ||| ||||||||| | |||| |||| |||||| | | ||
Db        301 ATTTCAACGNTGATGGATGGNCTTTAGAAGTACTTGGCTAGGCTCANTTGTGGGTNTNCA 360

Qy        361 TCATCATGCAAT---GGTGGGAAAGGGTCATTTTGGGGCCACGAGTGGGAGAAACATGGG 417
              |||||| | |||    | || || |||  ||||| ||| | ||  | || | ||    ||
Db        361 TCATCANGNAATTGGTGGGGAAAGGGGCAATTTTNGGGNCNCGGTTNGGGGNAANNAAGG 420

Qy        418 ACTTGTTCTTCTCCTGTTTTTCA 440
              | |||   |||  ||||| |
Db        421 GGCTTTTCCGGTCCCCTTTTTTA 443


Search completed: January 22, 2004, 00:54:32
Job time : 2026.55 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 16:52:57 ; Search time 68.455 Seconds (without alignments) 6525.161 Million cell updates/sec

Title: US-09-936-869-5
Perfect score: 1012
Sequence: 1 atcgaattaaagtcaatggc..........ttctatggtaaaaaaaaaaa 1012

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	170.8	16.9	822	5	PCT-US94-00844-10	Sequence 10, Appl
	2	39.4	3.9	1278	2	US-08-986-963-1	Sequence 1, Appli
	3	39.4	3.9	1278	2	US-08-986-963-3	Sequence 3, Appli
	4	39.4	3.9	2764	2	US-08-986-963-4	Sequence 4, Appli
c	5	39.4	3.9	13121	4	US-08-961-527-126	Sequence 126, App
c	6	38.2	3.8	3412	3	US-08-903-325-3	Sequence 3, Appli
c	7	38	3.8	786	3	US-08-948-265-3	Sequence 3, Appli
c	8	38	3.8	789	3	US-08-948-265-1	Sequence 1, Appli
	9	38	3.8	1342	4	US-09-489-847-89	Sequence 89, Appl
	10	37.8	3.7	831	3	US-09-007-484-3	Sequence 3, Appli
	11	37.8	3.7	831	3	US-09-309-682-3	Sequence 3, Appli
	12	37.8	3.7	1293	3	US-09-007-484-1	Sequence 1, Appli
	13	37.8	3.7	1293	3	US-09-309-682-1	Sequence 1, Appli
	14	37.8	3.7	3496	4	US-09-411-449-1	Sequence 1, Appli
	15	37.8	3.7	3916	4	US-09-411-449-2	Sequence 2, Appli
	16	37.8	3.7	6060	5	PCT-US96-09430-7	Sequence 7, Appli
	17	37.4	3.7	2168	3	US-08-749-522-6	Sequence 6, Appli
	18	37.2	3.7	5652	4	US-09-601-198-75	Sequence 75, Appl
	19	36.4	3.6	864	4	US-09-325-932A-45	Sequence 45, Appl
	20	36.4	3.6	1989	4	US-09-134-001C-1237	Sequence 1237, Ap
c	21	35	3.5	1887	4	US-09-601-198-4	Sequence 4, Appli
	22	35	3.5	4673	1	US-07-638-431-1	Sequence 1, Appli
	23	35	3.5	4673	5	PCT-US92-00018-1	Sequence 1, Appli
c	24	35	3.5	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	25	34.8	3.4	2907	4	US-09-107-532A-2316	Sequence 2316, Ap
c	26	34.4	3.4	260	6	5514558-9	Patent No. 5514558
c	27	34.4	3.4	4228	4	US-09-336-447A-16	Sequence 16, Appl
	28	34.2	3.4	1606	4	US-09-601-198-27	Sequence 27, Appl
c	29	34.2	3.4	24358	4	US-09-392-812A-1	Sequence 1, Appli
	30	34	3.4	559	1	US-08-469-667-19	Sequence 19, Appl
	31	34	3.4	559	4	US-09-224-110-19	Sequence 19, Appl
	32	34	3.4	559	5	PCT-US95-07289-19	Sequence 19, Appl
	33	34	3.4	1251	3	US-09-007-476-1	Sequence 1, Appli
c	34	34	3.4	1850	3	US-08-617-860B-32	Sequence 32, Appl
c	35	34	3.4	4098	2	US-08-605-106-4	Sequence 4, Appli
	36	34	3.4	4524	2	US-08-845-998-7	Sequence 7, Appli
	37	34	3.4	4524	3	US-09-206-537-7	Sequence 7, Appli
	38	34	3.4	4524	3	US-09-430-854-7	Sequence 7, Appli
	39	33.8	3.3	1248	4	US-09-134-001C-1980	Sequence 1980, Ap
	40	33.4	3.3	3701	4	US-09-583-545-1	Sequence 1, Appli
	41	33.4	3.3	3701	4	US-09-617-594A-5	Sequence 5, Appli
	42	33.4	3.3	3706	2	US-08-566-398-59	Sequence 59, Appl
	43	33.4	3.3	3706	2	US-08-658-665-63	Sequence 63, Appl
	44	33.4	3.3	3706	3	US-08-796-101-27	Sequence 27, Appl
	45	33.4	3.3	3706	3	US-09-085-273-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
PCT-US94-00844-10
; Sequence 10, Application PC/TUS9400844
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Method of Controlling Plant Pathogenic Fungi
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00844
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010403
; FILING DATE: 29-JAN-1993
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10..801
PCT-US94-00844-10

Query Match 16.9%; Score 170.8; DB 5; Length 822;
Best Local Similarity 56.0%; Pred. No. 2.8e-39;
Matches 366; Conservative 0; Mismatches 282; Indels 6; Gaps 2;

```
Qy   29 TATGTCTTCTCCTTCTCGTTGCGTGTATCGCCGGAGCATTTGCCGGAGACGTCATCGAAC 88
Db   41 TATCAGTTCTGCTTACCGGATTGTTCTCCGTTGATGGGAGAGCAGTTGATCTCGTTTCAA 100

Qy   89 TCAATCGATCTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCT 148
Db  101 CTCTCACCGAACACACAAAGTTTGATTTCTTCACACTGGCTCTGCAATGGCCAGCAACTT 160

Qy  149 ATTGCCGTGGAACT---CGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATG 205
Db  161 TCTGCTCAACGCATGAGAACAAATGCTGTCCTGAAAATGGTTGTTGCCAAGGAGGAAATT 220

Qy  206 CTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTT 265
Db  221 CTCCACCAGGATTTACAATCCACGGACTATGGCCCGATTACAGTGATGGAACATGGCCAT 280
```

```
Qy        266 CATGTTGTTATCGATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTG 325
              ||||||||     | ||||  | | | ||  |    |||||||| | ||| | |||| | ||
Db        281 CATGTTGTCCAGGTTCTGCTTATGATGAAACAAAGATTTCACCATTGCTTGATGCTTTGC 340

Qy        326 AGAAGTACTGGCCTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGT 385
              | |  ||   ||| |  || ||||||      |  |  |||||| ||     || |  ||
Db        341 ACACATATATGCCGATACTAAGTTGTAACGGTATAGAATCATGCGATAACAAGACAAAGT 400

Qy        386 CATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATG 445
              | |||||||   |  | |||||  ||||||| || |||||  |    ||        ||
Db        401 CGTTTTGGGCTTATCAATGGGAACAACATGGTACATGTTCGACCTTAGTAACAGGCGACC 460

Qy        446 AGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCC 505
              | |||   || ||| | || |  ||| |  | || ||| || |||||||| ||||| || |
Db        461 AATATGGATATTTCTTGACAAGCCTTGAATTGTATTTGGAGTATAATGTGACGGAAGTTC 520

Qy        506 TTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAA 565
              | |   |||||||    |    |||| ||     | |  ||  |    |  |  || ||
Db        521 TATCAAAAGCTGGAATTAAACCTTCAAATTCAAAGACTTACTCATCTGCTGCCATTGTTT 580

Qy        566 CAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAA---GAGATGCAA 622
              |  | ||| | ||  ||    |  | || || |||| |||||||    | |||| ||
Db        581 CTACTATTGAAACTGCTTTCGGTGCACCACCGCAATTGGTCTGCAAAAATGGTGATGTAA 640

Qy        623 TCGATGAAATACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTG 676
              |  | ||| | ||  | || ||    || |    ||| | | || ||||||
Db        641 TTAAAGAAGTCCGGTTGTGTTTTACCAAGCACCACAAGATCCGAGAGTGTGTGG 694
```

RESULT 2
US-08-986-963-1
; Sequence 1, Application US/08986963
; Patent No. 5958730
; GENERAL INFORMATION:
; APPLICANT: Rosteck Jr., Paul R.
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: FtsY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,963
; FILING DATE: December 8, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1278

US-08-986-963-1

Query Match 3.9%; Score 39.4; DB 2; Length 1278;
Best Local Similarity 52.1%; Pred. No. 0.13;
Matches 88; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   ||||    | || | |||       | ||| |||||| | ||   ||||||
Db       1104 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 1163

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||   | | | |  ||||  |      || || || |      | ||   |||||||| ||
Db       1164 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 1223

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
              |    |    ||   || | |||| |    |    ||  | |||| |
Db       1224 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGTTTA 1272
```

RESULT 3
US-08-986-963-3
; Sequence 3, Application US/08986963
; Patent No. 5958730
; GENERAL INFORMATION:
; APPLICANT: Rosteck Jr., Paul R.
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: FtsY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,963
; FILING DATE: December 8, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-986-963-3

Query Match 3.9%; Score 39.4; DB 2; Length 1278;
Best Local Similarity 35.5%; Pred. No. 0.13;
Matches 60; Conservative 28; Mismatches 81; Indels 0; Gaps 0;

```
Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   :|::    | || | :||      : |:| ||||||: : |:    |||||:
Db       1104 AACGGGAAUUGUUUUGACUAAGAUUGAUGGAACUGCUCGAGGAGGUGUGGUUCUAGCCAU 1163

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              :|   | | :  |:|:  |      || :| :: |      | ||   ||:|||:|| |:
Db       1164 UCGUGAAGAACUCAAUAUUCCUGUAAAAUUGAUUGGUUUUGGUGAAAAAAUCGAUGAUAU 1223
```

```
Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
              |    :    |:   || | :::| |    |   |:  |  ||:: |
Db       1224 UGGAGAGUUUAACUCAGAAAACUUUAUGAAAGGUCUCUUGGAAGGUUUA 1272


RESULT 4
US-08-986-963-4
; Sequence 4, Application US/08986963
; Patent No. 5958730
;  GENERAL INFORMATION:
;    APPLICANT:  Rosteck Jr., Paul R.
;    TITLE OF INVENTION:  Streptococcus Pneumoniae Gene Sequence
;    TITLE OF INVENTION:  FtsY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  US
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,963
;      FILING DATE:  December 8, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Webster, Thomas D.
;      REGISTRATION NUMBER:  39,872
;      REFERENCE/DOCKET NUMBER:  X-11763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317/276-3334
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2764 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-986-963-4

  Query Match             3.9%;  Score 39.4;  DB 2;  Length 2764;
  Best Local Similarity   52.1%;  Pred. No. 0.18;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   ||||    | || | |||       | ||| |||||||| | ||   ||||||
Db       1450 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 1509

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||   | | | |  ||||  |      || || || |      | ||   |||||||| ||
Db       1510 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 1569

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
              |    |   ||   || | |||| |    |   ||  | ||||  |
Db       1570 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGTTTA 1618


RESULT 5
US-08-961-527-126/c
; Sequence 126, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  126:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13121 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-126

  Query Match             3.9%;  Score 39.4;  DB 4;  Length 13121;
  Best Local Similarity   52.1%;  Pred. No. 0.34;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   ||||    | || | |||       | ||| |||||||| | ||   ||||||
Db       4788 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 4729

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||   | | | |  ||||  |      || || || |      | ||   |||||||| ||
Db       4728 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 4669

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
              |    |   ||   || | |||| |    |   ||  | ||||  |
Db       4668 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGTTTA 4620


RESULT 6
US-08-903-325-3/c
; Sequence 3, Application US/08903325
; Patent No. 6262338
;  GENERAL INFORMATION:
;    APPLICANT:  Peter Schreier; Thomas Herget; and
;    APPLICANT:  Jeff Schell
;    TITLE OF INVENTION:  RESISTANCE GENES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate SX/20
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/903,325
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/383,747
;       FILING DATE:  27-JAN-1995
;       APPLICATION NUMBER:  DE 40 31 758.7
;       FILING DATE:  06-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/766990
;       FILING DATE:  27-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/235106
;       FILING DATE:  28-APR-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  Bayer 8297.2-KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3412 nucleotides
;       TYPE:  Nucleic acid
;       STRANDEDNESS:  Double
;       TOPOLOGY:  Linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Peanut (arachis hypogaea)
US-08-903-325-3

  Query Match              3.8%;  Score 38.2;  DB 3;  Length 3412;
  Best Local Similarity   49.3%;  Pred. No. 0.42;
  Matches  100;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        756 TATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTT 815

Db       2386 TTTTGTCATGAATTAAAAAATAAATATATGAGGATTTTAATTTTGAGACACTAATAATGT 2327

Qy        816 TGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAAC 875

Db       2326 AAATAGATTATTTGATAATATTTATATATTTAGATAATTTTTTAATTTTAATATACTAAC 2267

Qy        876 CTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAG 935

Db       2266 AATGTAAAAATTAATTATTTCTATTGATATAACAATACACAATTAATTATGTGTAAAAAA 2207

Qy        936 TTTCTTTCTAATAATGTAGAAGA 958

Db       2206 TTTTATACTAATAATATATAAAA 2184

RESULT 7
US-08-948-265-3/c
; Sequence 3, Application US/08948265
; Patent No. 6228619
;  GENERAL INFORMATION:
;    APPLICANT:  Foster, Simon J.
;    APPLICANT:  Burnham, Martin K. R.
;    TITLE OF INVENTION:  No. 6228619el glucosaminidase
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert, Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2793
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,265
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/899008
;      FILING DATE:  23-JUL-1996
;      APPLICATION NUMBER:  US 60/011888
;      FILING DATE:  20-FEB-1996
;      APPLICATION NUMBER:  PCTUS97/02547
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dickinson, Q. Todd
;      REGISTRATION NUMBER:  28,354
;      REFERENCE/DOCKET NUMBER:  P50444-06
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215/994-2252
;      TELEFAX:  215/994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  786 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-948-265-3

  Query Match              3.8%;  Score 38;  DB 3;  Length 786;
  Best Local Similarity   49.0%;  Pred. No. 0.26;
  Matches  101;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        807 TGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTAC 866

Db        229 TAGCAATTTCATTATTATTTGCTTCAACAAATTTACCATCCTCTTCTTTTGTCTGTACAA 170

Qy        867 TACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGG 926

Db        169 TACCTTTCCCTTGTTGCATTGAAACAGCCTCATCAAAAGAGTAATTCACATCATTTTTAA 110

Qy        927 TAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACA 986

Db        109 ACAATTTAGTTTCATTCACGATGAGTAATACAGCAAAAACAACTAAAATAACTATCAATA 50

Qy        987 TTTGAGTTCTATGGTAAAAAAAAAAA 1012

Db         49 GTAGCGTTGAAATGCGTAACTTGAAA 24

RESULT 8
US-08-948-265-1/c
; Sequence 1, Application US/08948265
; Patent No. 6228619
;  GENERAL INFORMATION:
;    APPLICANT:  Foster, Simon J.
;    APPLICANT:  Burnham, Martin K. R.
;    TITLE OF INVENTION:  No. 6228619el glucosaminidase
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert, Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2793
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,265
```

```
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/899008
;      FILING DATE:  23-JUL-1996
;      APPLICATION NUMBER:  US 60/011888
;      FILING DATE:  20-FEB-1996
;      APPLICATION NUMBER:  PCTUS97/02547
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dickinson, Q. Todd
;      REGISTRATION NUMBER:  28,354
;      REFERENCE/DOCKET NUMBER:  P50444-06
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215/994-2252
;      TELEFAX:  215/994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  789 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-948-265-1

  Query Match             3.8%;  Score 38;  DB 3;  Length 789;
  Best Local Similarity   49.0%;  Pred. No. 0.26;
  Matches  101;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        807 TGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTAC 866
              | | | || |||| | ||   | | | |||   |||| | |||| |  | |
Db        229 TAGCAATTTCATTATTATTTGCTTCAACAAATTTACCATCCTCTTCTTTTGTCTGTACAA 170

Qy        867 TACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGG 926
              ||||    |||       |||| |      || | | || | |||||||  |
Db        169 TACCTTTCCCTTGTTGCATTGAAACAGCCTCATCAAAAGAGTAATTCACATCATTTTTAA 110

Qy        927 TAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACA 986
                |||  |||||| |||       |||  | | || |||  |  || |  ||||||  |
Db        109 ACAATTTAGTTTCATTCACGATGAGTAATACAGCAAAAACAACTAAAATAACTATCAATA 50

Qy        987 TTTGAGTTCTATGGTAAAAAAAAAAA 1012
              | | ||| |  |   ||   |||
Db         49 GTAGCGTTGAAATGCGTAACTTGAAA 24


RESULT 9
US-09-489-847-89
; Sequence 89, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: 98 Human Secreted Proteins
;  FILE REFERENCE: PZ031P1
;  CURRENT APPLICATION NUMBER: US/09/489,847
;  CURRENT FILING DATE: 2000-01-24
;  EARLIER APPLICATION NUMBER: PCT/US99/17130
;  EARLIER FILING DATE: 1999-07-29
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  EARLIER APPLICATION NUMBER: 60/095,486
;  EARLIER FILING DATE: 1998-08-05
;  EARLIER APPLICATION NUMBER: 60/096,319
;  EARLIER FILING DATE: 1998-08-12
;  EARLIER APPLICATION NUMBER: 60/095,454
;  EARLIER FILING DATE: 1998-08-06
;  EARLIER APPLICATION NUMBER: 60/095,455
;  EARLIER FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 376
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 89
;   LENGTH: 1342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-489-847-89

  Query Match             3.8%;  Score 38;  DB 4;  Length 1342;
  Best Local Similarity   54.2%;  Pred. No. 0.32;
  Matches   77;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        871 AAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAA 930
              |||| ||| ||  || |  | || ||     | | |  |  | ||  ||| ||||
Db       1183 AAAAGCTACCAATCTTTGTACAATTTGTAAATGTTAAGAATTTTTTTTATATCTGTTAAA 1242

Qy        931 TCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTG 990
              | ||  || |||| || ||   | || |  | ||||   | || || | |  || |
Db       1243 TAAAAATTATTTCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1302

Qy        991 AGTTCTATGGTAAAAAAAAAA 1012
              |     |  ||||||||||
Db       1303 AAAAAAAAAAAAAAAAAAAAA 1324


RESULT 10
US-09-007-484-3
; Sequence 3, Application US/09007484
; Patent No. 6072032
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael T.
;    TITLE OF INVENTION:  No. 6072032el FtsY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/007,484
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM10081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  831 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-007-484-3

  Query Match             3.7%;  Score 37.8;  DB 3;  Length 831;
  Best Local Similarity   51.5%;  Pred. No. 0.3;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||  ||||   | || | |||      | ||| |||||||| | ||  ||||||
Db        660 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 719

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||   | | | |  ||||  |     || || || |     | ||   |||||||| ||
```

```
Db        720 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 779

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681

Db        780 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGCTTA 828


RESULT 11
US-09-309-682-3
; Sequence 3, Application US/09309682
; Patent No. 6214348
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael T.
;    TITLE OF INVENTION:  No. 6214348el FtsY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/309,682
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/007,484
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM10081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  831 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-309-682-3

  Query Match             3.7%;  Score 37.8;  DB 3;  Length 831;
  Best Local Similarity   51.5%;  Pred. No. 0.3;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572

Db        660 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 719

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632

Db        720 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 779

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681

Db        780 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGCTTA 828


RESULT 12
US-09-007-484-1
; Sequence 1, Application US/09007484
; Patent No. 6072032
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael T.
;    TITLE OF INVENTION:  No. 6072032el FtsY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/007,484
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM10081
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1293 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-007-484-1

  Query Match             3.7%;  Score 37.8;  DB 3;  Length 1293;
  Best Local Similarity   51.5%;  Pred. No. 0.36;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572

Db       1119 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 1178

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632

Db       1179 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 1238

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681

Db       1239 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGCTTA 1287


RESULT 13
US-09-309-682-1
; Sequence 1, Application US/09309682
; Patent No. 6214348
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael T.
;    TITLE OF INVENTION:  No. 6214348el FtsY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
```

```
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/309,682
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  09/007,484
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Falk, Stephen T
;       REGISTRATION NUMBER:  36,795
;       REFERENCE/DOCKET NUMBER:  GM10081
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-994-2488
;       TELEFAX:  215-994-2222
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1293 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-309-682-1

  Query Match             3.7%;  Score 37.8;  DB 3;  Length 1293;
  Best Local Similarity   51.5%;  Pred. No. 0.36;
  Matches   87;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   |||| |  || | |||     | ||| ||||||| | ||   ||||||
Db       1119 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 1178

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||   | | | |  |||| |     || || || |  || |||||||| ||
Db       1179 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 1238

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
                |    |   ||  || | |||| |    |   ||  |  || | |
Db       1239 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGCTTA 1287


RESULT 14
US-09-411-449-1
; Sequence 1, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
;  APPLICANT:  James Ellis
;  TITLE OF INVENTION:  HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
;  TITLE OF INVENTION:  beta-GLOBIN REGULATORY ELEMENTS
;  FILE REFERENCE:  17860017
;  CURRENT APPLICATION NUMBER: US/09/411,449
;  CURRENT FILING DATE:  1999-10-01
;  PRIOR APPLICATION NUMBER:  2,246,005
;  PRIOR FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 3496
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-411-449-1

  Query Match             3.7%;  Score 37.8;  DB 4;  Length 3496;
  Best Local Similarity   48.8%;  Pred. No. 0.56;
  Matches  102;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        768 GATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTG 827
              |||  |  ||   |  ||   | |||||  | | ||  ||        ||| ||    | |
Db        915 GATTACCCCATTGATAGTCACACTTTGGGTTGTAAGTGACTTTTTATTTATTTGTATTTT 974

Qy        828 AGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTG 887
              ||    ||  | |  | |||    | || |  |||    | ||||||||||  |||
Db        975 TGACTGCATTAAGAGGTCTCTAGTTTTTTATCTCTTGTTTCCCAAAACCTAATAAGTAAC 1034

Qy        888 GCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAAT 947
              |    |  |   | |||  |||||       || |  || |   | ||| ||   |
Db       1035 TAATGCACAGAGCACATTGATTTGTATTTATTCTATTTTTAGACATAATTTATTAGCATG 1094

Qy        948 AATGTAGAAGATCAGAAAATCCATAAGAA 976
              |||   ||  | ||||||  | |  || ||
Db       1095 CATGAGCAAATTAAGAAAAACAACAACAA 1123


RESULT 15
US-09-411-449-2
; Sequence 2, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
;  APPLICANT:  James Ellis
;  TITLE OF INVENTION:  HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
;  TITLE OF INVENTION:  beta-GLOBIN REGULATORY ELEMENTS
;  FILE REFERENCE:  17860017
;  CURRENT APPLICATION NUMBER: US/09/411,449
;  CURRENT FILING DATE:  1999-10-01
;  PRIOR APPLICATION NUMBER:  2,246,005
;  PRIOR FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 3916
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-411-449-2

  Query Match             3.7%;  Score 37.8;  DB 4;  Length 3916;
  Best Local Similarity   48.8%;  Pred. No. 0.58;
  Matches  102;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        768 GATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTG 827
              |||  |  ||   |  ||   | |||||  | | ||  ||        ||| ||    | |
Db        915 GATTACCCCATTGATAGTCACACTTTGGGTTGTAAGTGACTTTTTATTTATTTGTATTTT 974

Qy        828 AGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTG 887
              ||    ||  | |  | |||    | || |  |||    | ||||||||||  |||
Db        975 TGACTGCATTAAGAGGTCTCTAGTTTTTTATCTCTTGTTTCCCAAAACCTAATAAGTAAC 1034

Qy        888 GCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAAT 947
              |    |  |   | |||  |||||       || |  || |   | ||| ||   |
Db       1035 TAATGCACAGAGCACATTGATTTGTATTTATTCTATTTTTAGACATAATTTATTAGCATG 1094

Qy        948 AATGTAGAAGATCAGAAAATCCATAAGAA 976
              |||   ||  | ||||||  | |  || ||
Db       1095 CATGAGCAAATTAAGAAAAACAACAACAA 1123


Search completed: January 21, 2004, 22:19:50
Job time : 71.455 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 21, 2004, 17:11:45 ; Search time 340.967 Seconds
(without alignments)
10461.606 Million cell updates/sec

Title: US-09-936-869-5
Perfect score: 1012
Sequence: 1 atcgaattaaagtcaatggc.....ttctatggtaaaaaaaaaaa 1012

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	208.4	20.6	1145	12	US-10-310-154-178	Sequence 178, App
	2	83.2	8.2	266	10	US-09-878-574-5684	Sequence 5684, Ap
	3	76.6	7.6	229	10	US-09-878-574-5737	Sequence 5737, Ap
	4	63.8	6.3	224	10	US-09-878-574-8638	Sequence 8638, Ap
	5	46.8	4.6	693	10	US-09-938-842A-1871	Sequence 1871, Ap
c	6	43.4	4.3	628	13	US-10-027-632-84237	Sequence 84237, A
c	7	43.4	4.3	628	13	US-10-027-632-296997	Sequence 296997,
c	8	43.4	4.3	628	14	US-10-027-632-84237	Sequence 84237, A
c	9	43.4	4.3	628	14	US-10-027-632-296997	Sequence 296997,
	10	40.8	4.0	19513	13	US-10-204-708-39	Sequence 39, Appl
c	11	40.4	4.0	10133	13	US-10-311-455-431	Sequence 431, App
	12	39.6	3.9	3673778	13	US-10-312-841-1	Sequence 1, Appli
	13	39.4	3.9	1278	9	US-09-815-242-9165	Sequence 9165, Ap
	14	39.2	3.9	381	10	US-09-878-574-1390	Sequence 1390, Ap
	15	39.2	3.9	387	10	US-09-878-574-953	Sequence 953, App
	16	39.2	3.9	598	13	US-10-027-632-199664	Sequence 199664,
	17	39.2	3.9	598	13	US-10-027-632-199665	Sequence 199665,
	18	39.2	3.9	598	14	US-10-027-632-199664	Sequence 199664,
	19	39.2	3.9	598	14	US-10-027-632-199665	Sequence 199665,
c	20	39	3.9	517	13	US-10-027-632-179698	Sequence 179698,
c	21	39	3.9	517	14	US-10-027-632-179698	Sequence 179698,
c	22	39	3.9	7810	13	US-10-240-453-155	Sequence 155, App
c	23	39	3.9	7810	15	US-10-239-676-141	Sequence 141, App
c	24	38.6	3.8	597	13	US-09-814-353-4872	Sequence 4872, Ap
c	25	38.6	3.8	597	13	US-09-814-353-11169	Sequence 11169, A
c	26	38.4	3.8	5520	13	US-10-311-455-1785	Sequence 1785, Ap
	27	38.4	3.8	6106	13	US-10-311-455-1446	Sequence 1446, Ap
	28	38.4	3.8	6123	13	US-10-311-455-793	Sequence 793, App
	29	38.4	3.8	6265	13	US-10-311-455-1390	Sequence 1390, Ap
c	30	38.2	3.8	3673778	13	US-10-312-841-1	Sequence 1, Appli
c	31	38	3.8	786	9	US-09-770-075-3	Sequence 3, Appli
c	32	38	3.8	786	9	US-09-769-997-3	Sequence 3, Appli
c	33	38	3.8	789	9	US-09-770-075-1	Sequence 1, Appli
c	34	38	3.8	789	9	US-09-769-997-1	Sequence 1, Appli
	35	38	3.8	1406	13	US-10-027-632-124113	Sequence 124113,
	36	38	3.8	1406	14	US-10-027-632-124113	Sequence 124113,
	37	37.8	3.7	831	9	US-09-827-663-3	Sequence 3, Appli
	38	37.8	3.7	1290	9	US-09-815-242-9491	Sequence 9491, Ap
	39	37.8	3.7	1293	9	US-09-827-663-1	Sequence 1, Appli
	40	37.8	3.7	1932	13	US-10-027-632-97574	Sequence 97574, A
c	41	37.8	3.7	1932	13	US-10-027-632-100117	Sequence 100117,
	42	37.8	3.7	1932	14	US-10-027-632-97574	Sequence 97574, A
c	43	37.8	3.7	1932	14	US-10-027-632-100117	Sequence 100117,
	44	37.8	3.7	15592	13	US-10-311-455-1300	Sequence 1300, Ap
	45	37.8	3.7	73308	10	US-09-954-456-2276	Sequence 2276, Ap

ALIGNMENTS

RESULT 1
US-10-310-154-178
; Sequence 178, Application US/10310154
; Publication No. US20030233670A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Chomet, Paul S.
; APPLICANT: Adams, Thomas H
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Agarwal, Ameeta K.
; APPLICANT: Ahrens, Jeffrey E.
; APPLICANT: Ball, James A.
; APPLICANT: Banu, G.
; APPLICANT: Bell, Erin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Deng, Molian
; APPLICANT: Dong, Jinzhuo
; APPLICANT: Duff, Stephen M.
; APPLICANT: Galligan, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shihshieh
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Lin, Jie-Yi
; APPLICANT: Liu, Jingdong
; APPLICANT: Lu, Bin
; APPLICANT: Luethy, Michael M.
; APPLICANT: Lund, Adrian
; APPLICANT: Madson, Linda L.
; APPLICANT: Malloy, Kathleen A.
; APPLICANT: McKiel, Christine L.
; APPLICANT: Miller, Philip W.
; APPLICANT: Padmavathi, Manchikanti

; APPLICANT: Parnell, Laurence D.
; APPLICANT: Start, William G.
; APPLICANT: Tennesen, Dan
; APPLICANT: Vidya, K.R.
; APPLICANT: Wang, Haiyun
; APPLICANT: Xin, Zhanguo
; APPLICANT: Xu, Nanfei
; APPLICANT: Yang, Chunzhi
; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, Yajuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310,154
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; PRIOR FILING DATE: 2001-12-04
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 178
; LENGTH: 1145
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (55)..(888)
; OTHER INFORMATION:
US-10-310-154-178

Query Match 20.6%; Score 208.4; DB 12; Length 1145;
Best Local Similarity 61.2%; Pred. No. 2.3e-49;
Matches 354; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

```
Qy         98 CTCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTG 157
              ||| |||||||||| || ||||| || || || || || ||||| || |||    |||
Db        158 CGCAGCGGGAGTTCGACTACTTCGCGCTCTCCCTCCAGTGGCCCGGCACCATCTGCGCCT 217

Qy        158 GAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAAT 217
                || ||||| || ||| |||  ||||  |||||| |   ||| ||  | |         |
Db        218 CCACCCGCCACTGCTGCGCCACCAACGGCTGCTGCCG---CTCGGAGCCGCTCCAGACGT 274

Qy        218 TCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTTCATGTTGTTATC 277
              |||| || || ||| ||||||| |||||  |||| ||  | ||||| || || ||    |
Db        275 TCACGATCCACGGGCTATGGCCGGACTACGACGACGGGACCTGGCCGTCGTGCTGCCGCC 334

Qy        278 GATCTGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTCTTGAGAAGTACTGGC 337
              | |  | ||  |   | |  |||| |   |  ||| ||| |  ||||| ||||||||||
Db        335 GCACCCAATTCGAGATGGACAAGATATTGCCACTGAAGGAGGTGCTTGACAAGTACTGGC 394

Qy        338 CTAGTCTCAGTTGTGGTTCTCCATCATCATGCAATGGTGGGAAAGGGTCATTTTGGGGCC 397
              |   |||   ||       |  | |||||   | ||| || ||    || ||||  |
Db        395 CGTCCCTCTACTGCTCCAAATCTGGAACATGCTTCAGCGGCAAGGGACTCTTCTGGGCTC 454

Qy        398 ACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATGAGTATAATTACT 457
              |||||||||||||||||| || || ||| | ||||  |||| |||||  |   | || |
Db        455 ACGAGTGGGAGAAACATGGAACCTGCTCTGCCCCTGTGGTTCAGGATGAATTACAGTATT 514

Qy        458 TCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCCTTTATCAAGCTG 517
              ||        | ||| | |||||||  ||| |||| || |||||  | ||| |      | |
Db        515 TCACCCTTGCCCTTGACCTCTACTCCAAGTATAACGTTACGGAAATGCTGTCAAGTGGAG 574

Qy        518 GCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCATTCAGA 577
              |       | ||| ||  || | | ||| | ||  | | | || |     |||| |
Db        575 GGATACAGGTTTCAAATGGTAAGGAATATGCACTGAGCGATGTCATCGATACCATCAAAC 634

Qy        578 ATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAATACGTA 637
              |||| |||      | ||  || | |||||||| || | | |||||| ||| | |
Db        635 ATGCTTTTGGGGGATCGCCACAAATTGTTTGCAAGAGGGGTTCAATCGAAGAACTTAGGT 694

Qy        638 TATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTT 675
              |||||||| | ||||| || |||||  | || ||  ||
Db        695 TATGCTTCGACAAAGAATTGAAGCCTCGTGATTGCCTT 732
```

RESULT 2
US-09-878-574-5684
; Sequence 5684, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 5684
; LENGTH: 266
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701097250H1
US-09-878-574-5684

Query Match 8.2%; Score 83.2; DB 10; Length 266;
Best Local Similarity 73.2%; Pred. No. 1.5e-13;
Matches 120; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

```
Qy         99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGG 158
              |||  | ||||| || |||||||||| |   |  | |||||||| || || || |||    |
Db        104 TCAACGAGAGTTTGACTATTTCGCTTTCGC-GTGCAATGGCCCGGTACTTACTGCAAACG 162

Qy        159 AACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATT 218
                 || |||  ||| |||  |||| |||||||||||||||| ||  ||||||    ||
Db        163 CACCCGCAGTTGCTGCCCCACCAACGGTTGCTGCAGAGGCTCCAATTTTCCAACAGTGTT 222

Qy        219 CACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGC 262
              ||| || |||||  | ||||||||||||||| |||||  | ||||
Db        223 CACCATACATGGACTCTGGCCTGACTATAATGATGGAACCTGGC 266
```

RESULT 3
US-09-878-574-5737
; Sequence 5737, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 5737
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701097310H1
US-09-878-574-5737

Query Match 7.6%; Score 76.6; DB 10; Length 229;
Best Local Similarity 67.5%; Pred. No. 1.1e-11;
Matches 137; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

```
Qy         99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTA-TTGCCGTG 157
              |||  | ||||| || ||| |||||  |   |  | |||||||| || || || | ||
Db         27 TCAACGAGAGTTTGACTATATCGCTTTCGCGTTGCAATGGCCCGGTACTTACTCGCAAAC 86

Qy        158 GAACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAAT 217
```

```
Db          87 GCACCCGCAGTTGCTGACCCACCAACGGTTGCTGCAGAAGCTCCAATTTTCCAACAGTGT 146

Qy         218 TCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCC-TTCATGTTGTTAT 276

Db         147 TCACCATACATGGACTCTGGCCTGACTATTATGATAGACCCTGGCCGTACCTGTTGCTCT 206

Qy         277 CGATCTGACTTTAAAGAGAAGGA 299

Db         207 GGATCTAGTTTCGATCCTAAAGA 229


RESULT 4
US-09-878-574-8638
; Sequence 8638, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 8638
;   LENGTH: 224
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701101468H1
US-09-878-574-8638

  Query Match             6.3%;  Score 63.8;  DB 10;  Length 224;
  Best Local Similarity   69.9%;  Pred. No. 5.6e-08;
  Matches   86;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          99 TCAGAGGGAGTTCGATTATTTCGCTCTATCTCTTCAATGGCCTGGAACCTATTGCCGTGG 158

Db         102 TCAACGAGAGTTTGACTATTTCGCTTTCGCGTTGCAATGGCCCGGTACTTACTGCAAACG 161

Qy         159 AACTCGCCATTGTTGCTCCAAAAACGCTTGCTGCAGAGGCTCCGATGCTCCAACTCAATT 218

Db         162 CACCCGCAGTTTCTGCCCCATCATCGGTTGCTGCAGAGGCTCCAATTTTCCAACAGTGTT 221

Qy         219 CAC 221

Db         222 CAC 224


RESULT 5
US-09-938-842A-1871
; Sequence 1871, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1871
;   LENGTH: 693
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-1871

  Query Match             4.6%;  Score 46.8;  DB 10;  Length 693;
  Best Local Similarity   56.0%;  Pred. No. 0.0094;
  Matches  131;  Conservative    0;  Mismatches   97;  Indels    6;  Gaps    2;

Qy         386 CATTTTGGGGCCACGAGTGGGAGAAACATGGGACTTGTTCTTCTCCTGTTTTTCATGATG 445

Db         341 CGTTTTGGGAGCACGAATGGGAGAAGCATGGTACTTGCTCTGAATCGGTTAT---CGATC 397

Qy         446 AGTATAATTACTTCCTTACCACACTTAATCTCTACTTGAAGCATAATGTCACGGATGTCC 505

Db         398 AACATGAATATTTCCAAACCGCTCTTAACCTTAAACAGAAAACCAATCTCCTTGGAGCTC 457

Qy         506 TTTATCAAGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAA 565

Db         458 TAACCAAAGCCGGGA---TTAATCCGGATGGAAAATCTTACTCTTTGGAGAGCATAAGAG 514

Qy         566 CAGCCATTCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATG 619

Db         515 ATTCGATAAAAGAGTCAATTGGTTTCACTCCTTGGGTTGAGTGTAACAGAGATG 568


RESULT 6
US-10-027-632-84237/c
; Sequence 84237, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 84237
;   LENGTH: 628
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-84237

  Query Match             4.3%;  Score 43.4;  DB 13;  Length 628;
  Best Local Similarity   49.8%;  Pred. No. 0.085;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         761 TTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTA 820

Db         375 TTGATAAAATACCAAAGGATACATTTTATATTCTAAAATGAAAAAGTATTCAAATGTTTA 316

Qy         821 TCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATA 880

Db         315 TCACCTGAAAAGTTATCTAAGCAGCTGTTTGTCTTGGAAATTTCAATAACAGAACACTGA 256

Qy         881 AAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCT 940
```

```
Db        255 AAACAATGTGCATAAAGTGCAAATAATAACAAGGCCACATTACTCACTAATGAATTATCT 196

Qy        941 TTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATAT 981

Db        195 TGATAATATTTAAATTCATGAAAAATAGTATATTAAGTCAT 155
```

RESULT 7
US-10-027-632-296997/c
; Sequence 296997, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 296997
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-296997

Query Match 4.3%; Score 43.4; DB 13; Length 628;
Best Local Similarity 49.8%; Pred. No. 0.085;
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        761 TTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTA 820

Db        375 TTGATAAAATACCAAAGGATACATTTTATATTCTAAAATGAAAAAGTATTCAAATGTTTA 316

Qy        821 TCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATA 880

Db        315 TCACCTGAAAAGTTATCTAAGCAGCTGTTTGTCTTGGAAATTTCAATAACAGAACACTGA 256

Qy        881 AAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCT 940

Db        255 AAACAATGTGCATAAAGTGCAAATAATAACAAGGCCACATTACTCACTAATGAATTATCT 196

Qy        941 TTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATAT 981

Db        195 TGATAATATTTAAATTCATGAAAAATAGTATATTAAGTCAT 155
```

RESULT 8
US-10-027-632-84237/c
; Sequence 84237, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 84237
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-84237

Query Match 4.3%; Score 43.4; DB 14; Length 628;
Best Local Similarity 49.8%; Pred. No. 0.085;
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy        761 TTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTA 820

Db        375 TTGATAAAATACCAAAGGATACATTTTATATTCTAAAATGAAAAAGTATTCAAATGTTTA 316

Qy        821 TCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATA 880

Db        315 TCACCTGAAAAGTTATCTAAGCAGCTGTTTGTCTTGGAAATTTCAATAACAGAACACTGA 256

Qy        881 AAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCT 940

Db        255 AAACAATGTGCATAAAGTGCAAATAATAACAAGGCCACATTACTCACTAATGAATTATCT 196

Qy        941 TTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATAT 981

Db        195 TGATAATATTTAAATTCATGAAAAATAGTATATTAAGTCAT 155
```

RESULT 9
US-10-027-632-296997/c
; Sequence 296997, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 296997
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-296997

```
  Query Match             4.3%;  Score 43.4;  DB 14;  Length 628;
  Best Local Similarity   49.8%;  Pred. No. 0.085;
  Matches  110;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        761 TTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGGAGCTTTGTTA 820
              ||   || || ||||  || | |     | ||  ||   |||||  |        | |||
Db        375 TTGATAAAATACCAAAGGATACATTTTATATTCTAAAATGAAAAAGTATTCAAATGTTTA 316

Qy        821 TCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATA 880
              ||  |||| |    ||  |  || | | ||   | | || || | || || |||    |
Db        315 TCACCTGAAAAGTTATCTAAGCAGCTGTTTGTCTTGGAAATTTCAATAACAGAACACTGA 256

Qy        881 AAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAATCAAGTTTCT 940
              ||      |   ||   || | | |     |     | |||||||    ||| || | |||
Db        255 AAACAATGTGCATAAAGTGCAAATAATAACAAGGCCACATTACTCACTAATGAATTATCT 196

Qy        941 TTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATAT 981
              |  |||||  |   |    || | |||   |||  |||  ||
Db        195 TGATAATATTTAAATTCATGAAAAATAGTATATTAAGTCAT 155


RESULT 10
US-10-204-708-39
; Sequence 39, Application US/10204708
; Publication No. US20030141852A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 39
;   LENGTH: 19513
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-39

  Query Match             4.0%;  Score 40.8;  DB 13;  Length 19513;
  Best Local Similarity   50.5%;  Pred. No. 4.3;
  Matches   99;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         789 TCTTTGAATCGGAAAAGATGGGAGCTTTGTTATCTTCTGAGAGACAATACATACATGTCT 848
               | ||  | |   | || |  |  | ||   ||| || ||    | |||| | |      |
Db        9107 TTTTGTATTATAATAAAAAAGACGTTTATATATTTTTTGTATTATAATAAAAAAGACGTT 9166

Qy         849 CTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGAT 908
               || || |   | |   |  |||   ||      ||  || ||| || | ||| |
Db        9167 TATATATTTTTTGTATTAAAAGTAAAAAGATTTTTATATATTTTTTTGTTTATTTTGTAA 9226

Qy         909 ATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATC 968
               || | | || |||  | |  ||  | ||| ||| ||||| | |  ||  |   ||| |
Db        9227 ATTTGTTATTATTTTTTATTTTATATATTTTTTTTTTTAATATTTTGAAAAGTTTTAAATTT 9286

Qy         969 CATAAGAAGATATCAA 984
               |||||  |||  ||
Db        9287 GTAGAGAAGGTATAAA 9302
```

```
RESULT 11
US-10-311-455-431/c
; Sequence 431, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Dete
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 431
;   LENGTH: 10133
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-431

  Query Match             4.0%;  Score 40.4;  DB 13;  Length 10133;
  Best Local Similarity   52.4%;  Pred. No. 3.7;
  Matches   89;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         843 ATGTCTCTGATGTTGTAACTTTACTACCAAAACCTATAAAGATTGGCTTATTTCGTTCTA 902
               ||  |  | | ||   |  | || |  ||||  |||| | ||     |  |     | |
Db        2924 ATAACAATTACGTACAATATATAATTACAAATTATATATATATATATTATAAAAACAA 2865

Qy         903 TTGGATATGTATCATCATTACTGGTAAATCAAGTTTCTTTCTAATAATGTAGAAGATCAG 962
               |  | |  ||| || | || |  |||  ||  | ||  ||  | ||   | ||    |
Db        2864 CTACAAAAATATAATAACTATTATAAAAAAAACATACTAACTTAAAACCAAAAATTCAAA 2805

Qy         963 AAAATCCATAAGAAGATATCAACATTTGAGTTCTATGGTAAAAAAAAAAA 1012
               | || ||  || || ||| ||| |  |  |  ||||    ||||||||||
Db        2804 ACAAACCTAAACAACATAACAAAACCTTATCTCTACTAAAAAAAAAAAAA 2755


RESULT 12
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MH
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             3.9%;  Score 39.6;  DB 13;  Length 3673778;
  Best Local Similarity   46.9%;  Pred. No. 2.4e+02;
  Matches  123;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;
```

```
Qy        751 GAAGCTATGGTTCTGAAGATGCCAACAGAAAGAGAAGCTCTTTGAATCGGAAAAGATGGG 810
              ||||  ||| ||  |||||     | | ||| ||        || |  |||||  ||| |
Db     863063 GAAGGGATGTTTTGGAAGAGAAGATAATAAATAGGTTTAAGTTTGAAGGGAAATTATGAG 863122

Qy        811 AGCTTTGTTATCTTCTGAGAGACAATACATACATGTCTCTGATGTTGTAACTTTACTACC 870
              || |||       ||| ||          | | |  | | | | | |
Db     863123 GTTGTTTTTAGTAGGGGAGGGAGTTGGATATGGTTTTTTATTTTAGGGAAATAAGGATG 863182

Qy        871 AAAACCTATAAAGATTGGCTTATTTCGTTCTATTGGATATGTATCATCATTACTGGTAAA 930
              | |     | || | |  ||  ||   | | ||     | |||    | |||  ||| ||
Db     863183 GAGAGGGGGAGAGTTAGAATTTGTTGAATTTGTTTTGGTTTTATTGGGAATATAGGTTAA 863242

Qy        931 TCAAGTTTCTTTCTAATAATGTAGAAGATCAGAAAATCCATAAGAAGATATCAACATTTG 990
              |  ||| |    | ||  | |||||     || | ||  ||| | | |||| |      ||
Db     863243 ATAGTTTTTTAGGTTATTTTTTAGAATGAAAGTATATAGATAGGTAAATATGATATAATG 863302

Qy        991 AGTTCTATGGTAAAAAAAAAA 1012
              | | | || || || |   ||
Db     863303 TGATTTGTGTTATAATATTTAA 863324
```

RESULT 13
US-09-815-242-9165
; Sequence 9165, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9165
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1278)
US-09-815-242-9165

Query Match 3.9%; Score 39.4; DB 9; Length 1278;
Best Local Similarity 52.1%; Pred. No. 1.9;
Matches 88; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy        513 AGCTGGCTATGTTGCTTCCAACAGTGAAAAGTATCCTCTAGGAGGTATCGTAACAGCCAT 572
              | | ||   ||||    | || | |||       ||| |||||||| | ||   ||||||
Db       1104 AACGGGAATTGTTTTGACTAAGATTGATGGAACTGCTCGAGGAGGTGTGGTTCTAGCCAT 1163

Qy        573 TCAGAATGCATTTCATATCACCCCTGAAGTGGTTTGCAAAAGAGATGCAATCGATGAAAT 632
              ||  | | | |||| |     || || || |     | ||   |||||||| ||
Db       1164 TCGTGAAGAACTCAATATTCCTGTAAAATTGATTGGTTTTGGTGAAAAAATCGATGATAT 1223

Qy        633 ACGTATATGCTTCTATAAAGATTTTAAGCCCAGGGACTGTGTTGGTTCA 681
              |     |   ||  | |||| |    |   ||  |  ||||  |
Db       1224 TGGAGAGTTTAACTCAGAAAACTTTATGAAAGGTCTCTTGGAAGGTTTA 1272
```

RESULT 14
US-09-878-574-1390
; Sequence 1390, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 1390
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-038-Q1-B1-F11
US-09-878-574-1390

Query Match 3.9%; Score 39.2; DB 10; Length 381;
Best Local Similarity 58.8%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 58; Indels 3; Gaps 1;

```
Qy        206 CTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTT 265
              || |  || | ||||  |||||||  | |||||| |||  || ||||| || ||||| |
Db        226 CTGCTGCTGATTTCAGCATTCATGGACTCTGGCCTAACTTCAAGGATGGCTCATGGCCCT 285

Qy        266 CATGTTGTTATCGATC---TGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTC 322
              ||   ||| | |      || |||| | |    |||| |||    | |||
Db        286 CAAACTGTGACCCTGACAGTGTCTTTGATAAATCTCAGATCTCAGACCTCATTAGCAACA 345

Qy        323 TTGAGAAGTACTGGCCTAGTCTCAGTTG 350
              | |||||| | |||||    || || ||
Db        346 TGGAGAAGGATTGGCCATCCCTGAGCTG 373
```

RESULT 15
US-09-878-574-953
; Sequence 953, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 953
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-043-Q1-B1-H8
US-09-878-574-953

Query Match 3.9%; Score 39.2; DB 10; Length 387;
Best Local Similarity 58.8%; Pred. No. 1;

Matches 87; Conservative 0; Mismatches 58; Indels 3; Gaps 1;

```
Qy        206 CTCCAACTCAATTCACAATTCATGGGTTATGGCCTGACTATAACGATGGTTCGTGGCCTT 265
              ||  |  ||  |  ||||  ||||||||  | ||||||| |||  || ||||| || ||||| |
Db        226 CTGCTGCTGATTTCAGCATTCATGGACTCTGGCCTAACTTCAAGGATGGCTCATGGCCCT 285

Qy        266 CATGTTGTTATCGATC---TGACTTTAAAGAGAAGGAGATTTCAACGTTGATGGATGGTC 322
              ||   |||  |  |      ||  ||||  |  |     ||||  |||    | ||
Db        286 CAAACTGTGACCCTGACAGTGTCTTTGATAAATCTCAGATCTCAGACCTCATTAGCAACA 345

Qy        323 TTGAGAAGTACTGGCCTAGTCTCAGTTG 350
              |  ||||||  |  |||||    || ||  ||
Db        346 TGGAGAAGGATTGGCCATCCCTGAGCTG 373
```

Search completed: January 22, 2004, 01:20:48
Job time : 349.967 secs